博士論文

Computational Study on the Catalytic Mechanism and Substrate Specificity of Enzymes from the Meroterpenoid Biosynthetic Pathway

(メロテルペノイドの生合成経路における酵素の触媒機構と基質特異性の研究)

趙　凡

博士論文

Computational Study on the Catalytic Mechanism and Substrate Specificity of Enzymes from the Meroterpenoid Biosynthetic Pathway
（メロテルペノイドの生合成経路における酵素の触媒機構と基質特異性の研究）

東京大学 大学院農学生命科学研究科 応用生命工学専攻 生物情報工学研究室

令和 3 年度　進学

氏名　　趙　凡

指導教員名　寺田　透

Contents

Abbreviations

8-AF	8-amino-flaviolin
8-DFA	8-diazoflaviolin
DFT	density functional theory
DMAPP	dimethylallyl diphosphate
DMATS	dimethylallyltryptophan synthase
FPP	farnesyl diphosphate
FLA	flaviolin
GAFF2	general AMBER force field 2
GGPP	geranylgeranyl diphosphate
GPP	geranyl diphosphate
HF	Hartree–Fock
HOMO	highest occupied molecular orbital
IM1	intermediate 1
IM2	intermediate 2
IPA	Invariant Point Attention
IPP	isopentenyl diphosphate
MD	molecular dynamics
MEP	methylerythritol phosphate
3-Me-PHN	3-methyl-1,2,4,5,7-pentahydroxy naphthalene
MM	molecular mechanics
MMF	2-methoxy-3-methyl-flaviolin
MMT	6-methoxy-7-methylnaphthalene-1,3,5,8-tetraol
MOP	mompain
MSA	multiple sequence alignment
MVA	mevalonate
ONIOM	our own n-layer integrated molecular orbital and molecular mechanics
ONIOM-EE	ONIOM electronic embedding
ORF	open reading frame

PCM	polarizable continuum model
PES	potential energy surface
PHN	1,2,4,5,7-pentahydroxynaphthalene
pLDDT	predicted local distance difference test
PME	particle mesh Ewald
PND	2,4,5,7,8-pentahydroxynaphthalene-1-diazonium
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
RESP	restrained electrostatic potential
SAH	*S*-adenosyl-L-homocysteine
SAM	*S*-adenosyl-L-methionine
SET	suppressor of variegation, enhancer of zeste and trithorax
THN	tetrahydronaphthalene

Abstract

Bacteria belonging to the *Streptomyces* genus have attracted significant attention for their ability to produce various secondary metabolites with complex molecular scaffolds and extensive bioactivities. Among them, the *Streptomyces* sp. strain KO-3988 produces furaquinocin D, a natural polyketide-isoprenoid hybrid that exhibits antitumor and cytotoxicity activities. In its biosynthetic process, a methyltransferase, Fur6, and a prenyltransferase, Fur7, respectively transfer the methyl and prenyl groups to the substrates to modulate their bioactivities. However, the studies on their catalytic mechanisms and substrate specificity are still insufficient because their substrates are easily oxidized. While over the past few decades, computational tools have demonstrated their utility in the mechanistic investigation of enzymes.

In this study, I report the catalytic mechanism and substrate specificity of Fur6 and Fur7 based on computational studies. For the study on Fur6, molecular dynamics (MD) simulation studies reveal the reactive form of the substrate, 1,2,4,5,7-pentahydroxynaphthalene (PHN), and its interactions with the enzyme. Hybrid quantum mechanics/molecular mechanics (QM/MM) calculations suggest the reaction pathway of the methyl transfer step, in which the energy barrier is 8.6 kcal mol^{-1}. Free-energy calculations with a polarizable continuum model (PCM) indicate that the final deprotonation step of the methylated intermediate occurs after it is ejected into the water solvent from the active center pocket of Fur6. Additionally, studies on the protonation states, the highest occupied molecular orbital (HOMOs), and the energy barriers of the methylation reaction for the analogs of PHN demonstrate the mechanism of the specificity to PHN. For the study on Fur7, analyses of the HOMOs reveal the original substrate of Fur7 should be 6-methoxy-7-methylnaphthalene-1,3,5,8-tetraol (MMT). MD simulation studies suggest the interactions among Fur7, MMT, and geranyl diphosphate (GPP). QM/MM calculations indicate the reactive form of MMT and the reaction pathway of the prenyl transfer step, in which the energy barrier is 0.42 kcal mol^{-1}.

Overall, this study provides valuable insights into Fur6 and Fur7 chemistry, not only contributing to a deeper understanding of molecular mechanisms but also offering an opportunity to better engineer the active site and control the reactivity to obtain high yields of the desired product(s).

1 Introduction

1.1 Background

1.1.1 Meroterpenoids are Potential Natural Secondary Metabolites

Throughout the course of history, dating back to ancient civilizations, natural products have held a significant position in the realm of disease treatment.[1] This is particularly evident in the areas of infectious diseases, cancer,[2,3] and a range of other ailments, including multiple sclerosis and cardiovascular conditions.[4-6] Even nowadays, natural products and their analogs continue to play a vital role in the development of novel drugs, primarily due to their remarkable scaffold diversity and intricate structural compositions.[7] According to previous reports, among the 877 New Chemical Entities introduced between 1981 and 2002, nearly 49% can be attributed to natural products or synthetic compounds that draw inspiration from natural product pharmacophores.[8,9]

In the realm of natural products, meroterpenoids represent a diverse class of secondary metabolites that find widespread occurrence in nature. These compounds are synthesized by a wide range of organisms, including unicellular ones like bacteria and fungi, as well as multicellular organisms like plants and animals.[10] Interestingly, the word meroterpenoid was initially introduced by Sir John W. Cornforth in an article focused on the biosynthesis of terpenoids.[11] The prefix "*mero*-" is derived from the Greek root "*merus*" which means "fragment, or part, or partial".[12] Meroterpenoids are a class of compounds that derived from the hybridization of polyketide and non-polyketide biosynthesis with terpenoid pathways (Figure 1.1).[13] Among the class of meroterpenoid, a significant number of them have been isolated from actinomycetes. For instance, naphthterpin was discovered in *Streptomyces* sp. CL190,[14] furaquinocin D was isolated from both *Streptomyces* sp. KO-3988 and *Streptomyces reveromyceticus* SN-593,[15] napyradiomycin A1 was extracted from *Streptomyces* sp. CNQ-525,[16] marinone was isolated from *Streptomyces* sp. CNQ-509,[17] merochlorin A was obtained from *Streptomyces* sp. CNH-189,[17] and marfuraquinocin A was isolated from *Streptomyces niveus* SCSIO 3406.[18]

The biosynthetic combination creates not only leads to the creation of a vast array of chemical diversity but also endows meroterpenoids with the potential for various bioactivities.[19] These compounds exhibit a range of beneficial effects, including antitumor,[20] antioxidant,[21,22] anti-inflammatory,[23,24] anti-diabetic,[25] and neuroprotective activities.[26] Over the past few decades, meroterpenoids have been widely studied. Geris and Simpson reported a comprehensive review of meroterpenoids in 2009,[13] El-Demerdash et al. focused on 320 meroterpenoids covering from

2009 to 2019,[27] while Zhao et al. summarized 709 meroterpenoids from 2009 to 2020.[28] Additionally, Nazir et al. described 452 new meroterpenoids from extensive natural sources that were discovered from 2016 to 2020.[29]

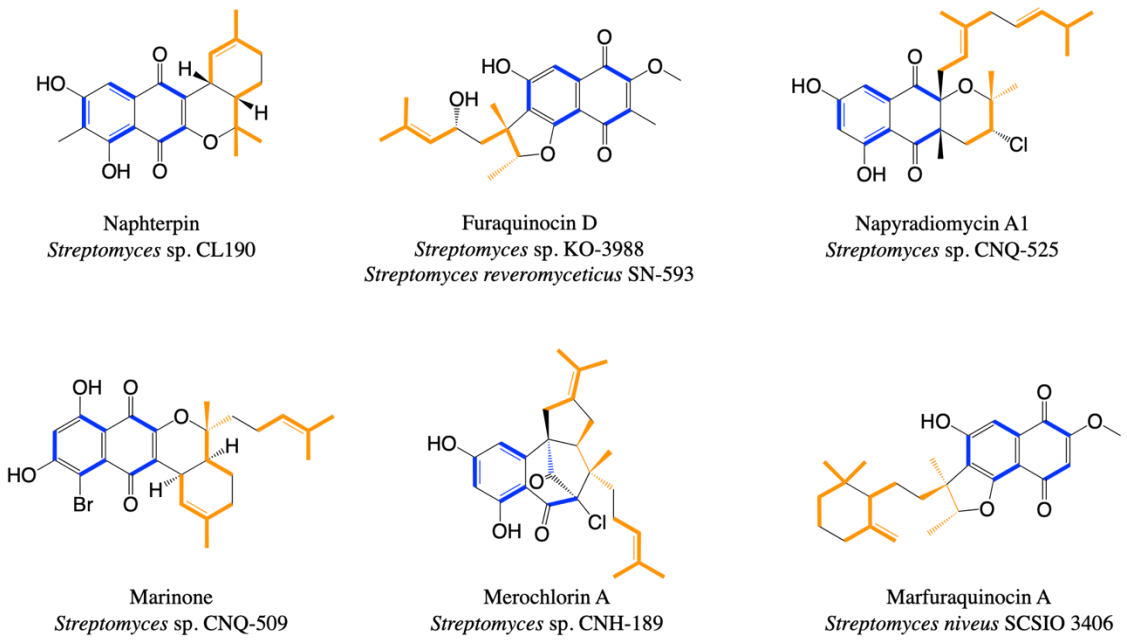


Figure 1.1 Presentative structures of meroterpenoids. The polyketide scaffolds are shown in blue, and the terpenoid scaffolds are shown in orange.

Furthermore, as genome exploration technologies have advanced significantly, researchers have started to unveil the intricate biosynthetic pathways of meroterpenoids (Figure 1.2). For instance, Kuzuyama, T. et al. elucidated the biosynthetic gene cluster of naphterpin in 2005,[30] followed by Kawasaki, T. et al. reported the biosynthetic gene cluster of furaquinocin in 2006,[31] and Winter, J. M. et al. demonstrated the biosynthetic gene cluster of Napyradiomycin A1 in 2007.[32] A fascinating revelation from these studies is that, despite the distinct structures of their final products, they share a common set of biosynthetic genes. The THN synthase genes (e.g., *fur1* for furaquinocin biosynthesis) are found in the biosynthetic gene clusters for each meroterpenoid. This commonality results in these compounds all having a tetrahydronaphthalene (THN)-derived skeleton. (Figure 1.1).

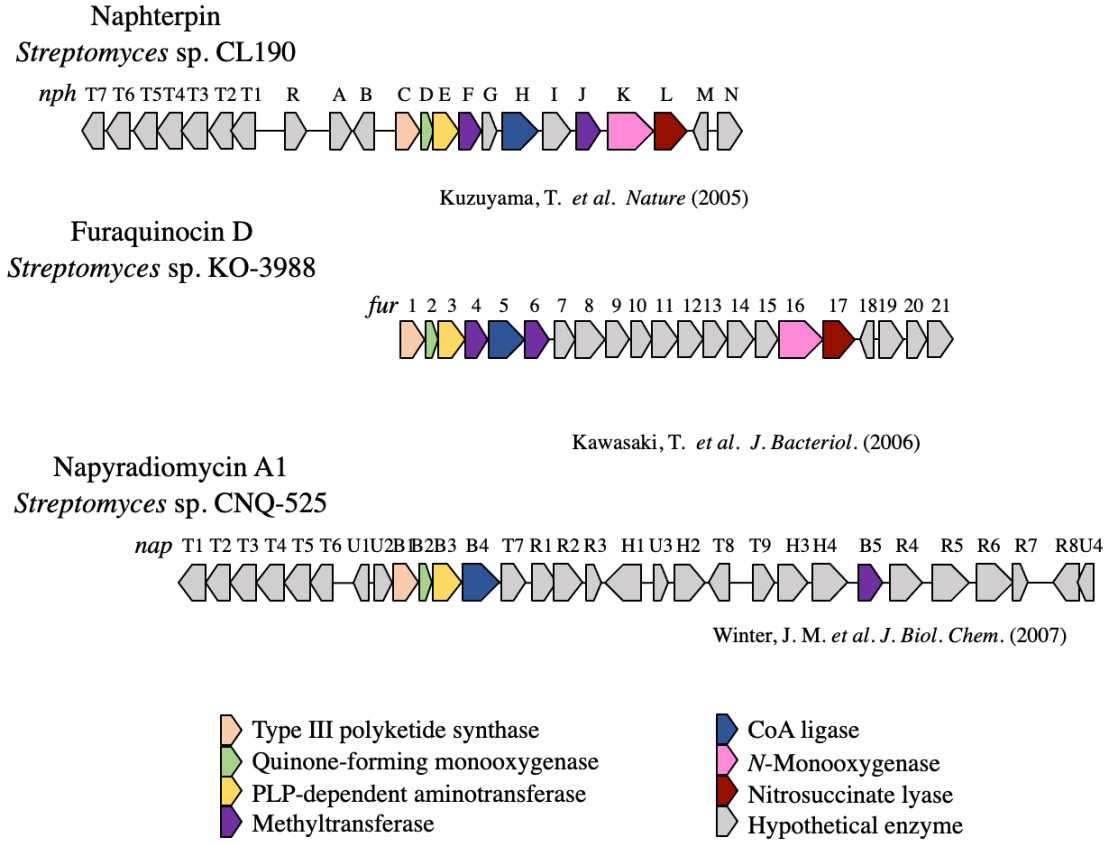


Figure 1.2 Gene clusters responsible for producing meroterpenoids. Crucial genes are conserved among individual meroterpenoid gene clusters.

In the class of meroterpenoids, furaquinocins, as briefly mentioned in the previous text, were initially discovered in 1989 and possess antitumor and cytocidal activities against HeLa S3 and B16 melanoma cells.[15] The biosynthetic gene cluster was deduced by Kawasaki, T. et al. in 2006, consisting of 21 open reading frames (ORF) (Table 1.1).[33] Since they estimated the functions of most of the genes, they also briefly proposed the biosynthetic pathway of the furaquinocins. Subsequently, the prenylation step catalyzed by the prenyltransferase Fur7 is investigated by Kumano, T. et al. in 2010.[20] Recently, the entire biosynthetic pathway for each step was proposed by Noguchi, T et al.[34] Comparing it to the proposed biosynthetic pathway from other researchers, Noguchi, T et al. deduced that from the reaction catalyzed by Fur6, the aunthetic substrate should be the naphthalene skeleton rather than the naphthalenone skeleton. The biosynthetic building blocks of furaquinocins consist of five molecules of malonyl-CoA, which subsequently undergo Claisen condensation reactions catalyzed by *Streptomyces* type III polyketide synthase Fur1.[35] The resulting THN skeleton is oxidized to flaviolin (FLA) then further oxidized to mompain (MOP) catalyzed by quinone forming monooxygenase Fur2. Then, MOP is further modified through a PLP-dependent aminotransferase Fur3, and 8-amino-flaviolin (8-AF) is produced. 8-AF is a common biosynthetic intermediate not only in furaquinocins but also in other meroterpenoids biosynthetic pathways. Through subsequent deamination, methylation, prenylation, and cyclization, the final furaquinocins are synthesized (Figure 1.3).[20,34] However,

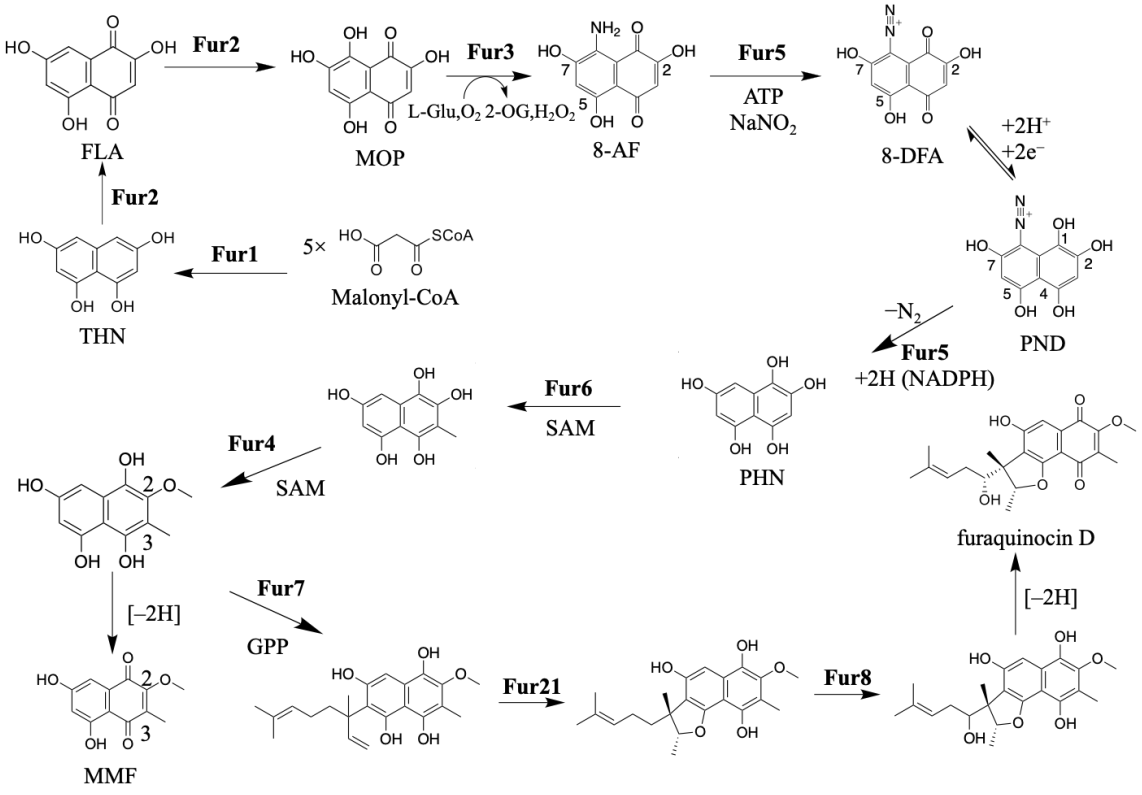


Figure 1.3 Proposed biosynthetic pathway of Furaquinocin D.

among these modifications, only prenylation catalyzed by Fur7 has been investigated. This indicates that, although general biosynthetic pathways of furaquinocins have been proposed, the precise reaction mechanisms involving the biosynthetic enzymes remain shrouded in mystery.

Table 1.1 *fur* gene cluster and homologous proteins of each gene

ORFs	Homologous proteins and their accession numbers
fur1	Type III polyketide synthase of *Streptomyces antibioticus* (88% identity), DAD; AB084489-1
fur2	Hypothetical protein of *Sacchropolyspora erythraea* (61%), DAD; AY078067-7
fur3	Aminotransferase of *Streptomyces griseus* (36%), DAD; Y00459-10
fur4	*C*-Methyl-transferase of *Streptomyces griseus* (28%), DAD; AJ578458-15
fur5	Fatty acid-CoA ligase of *Listeria innocua* (36%), DAD; AL596164-139
fur6	*O*-Methyl-transferase of *Photobacterium profundum* (25%), DAD; CR378675-122
fur7	Prenyltranferase (CloQ like protein) of *Streptomyces roseochromogenes* (23%), DAD; AF329398-29
fur8	P450 protein of *Streptomyces peucetius* (30%), DAD; AJ605552-1
MK	Mevalonate kinase of *Streptomyces* sp. CL190 (69%), DAD; AB037666-1
MDPD	Mevalonate diphosphate decarboxylase of *Streptomyces* sp. CL190 (77%), DAD; AB037666-2
PMK	Phosphomevalonate kinase of *Streptomyces* sp. CL.190 (69%), DAD; AB037666-3
IPP iso	Type 2 IPP isomerase of *Streptomyces* sp. CLI90 (79%), DAD; AB037666-4
HMGR	HMG-CoA reductase of *Streptomyces* sp. CL190 (88%), DAD; AB037666-5
HMGS	HMG-CoA synthase of *Streptomyces* sp. CL190 (79%), DAD; AB037666-6
fur15	3-oxoacyl-(acyl-carrier-protein) synthase of *Actinoplanes* sp. (60%), DAD; AB113568-8
fur16	Hypothetical protein of *Corynebacteriam glutamicum* (33%), DAD; AB193029-7
fur17	3-carboxy-cis, cis-muconate cycloisomerase of *Streptomyces coelicolor* (55%), DAD; AL939128-219
fur18	Hypothetical protein of *Flankia* sp. (57%), DAD; AY027524-1
fur19	Polyprenyl diphosphate synthase of *Mycobacterium avium* (36%), DAD; AE017234-5
fur20	Hypothetical protein of *Streptomyces coelicolor* (64%), DAD; AL939118-16
fur21	Methyl transferase of *Streptomyces ambofaciens* (26%), DAD; AY338477-1

1.1.2 Methyltransferases Play an Important Role in Molecular Modification

Methylation is a universal reaction in all living organisms and plays an important role in different biological processes such as cell signaling, membrane components, and pigments.[36] In general, methylation is an important step for the diversification of the structure of natural products in biosynthetic pathways, which can be utilized in chemo-, regio- and stereospecific synthesis. Methylation is also a common transformation in the biosynthesis of meroterpenoids, which helps modulate the bioactivity of the acceptor molecule.

Methyltransferases have a long history, over millions of years, they have greatly expanded and diversified a lot. Although their evolutionary picture becomes complex, they can be divided into five classes.[37]

Class I, also called the Rossmann-like fold family, is the major class with a seven-strand twisted β-sheet structure and sandwiched by helices on each side (Figure 1.4 A), which was first recognized in 1989.[38,39] The binding region of cofactor *S*-adenosyl-L-methionine (SAM) at the C-terminal ends of β1 and β2, the hydrogen bonding residue accommodates the ribose moiety, the acidic residue coordinates the methionine, while a conserved GxGxG motif interacts with the adenine moiety of SAM.[37] On the contrary, the substrate binding region always changes a lot, sometimes the additional domains are embedded within the Rossmann fold, or the additional domains are appended to the Rossmann fold.[40] These changes lead to the multiple substrate specificity of methyltransferases.

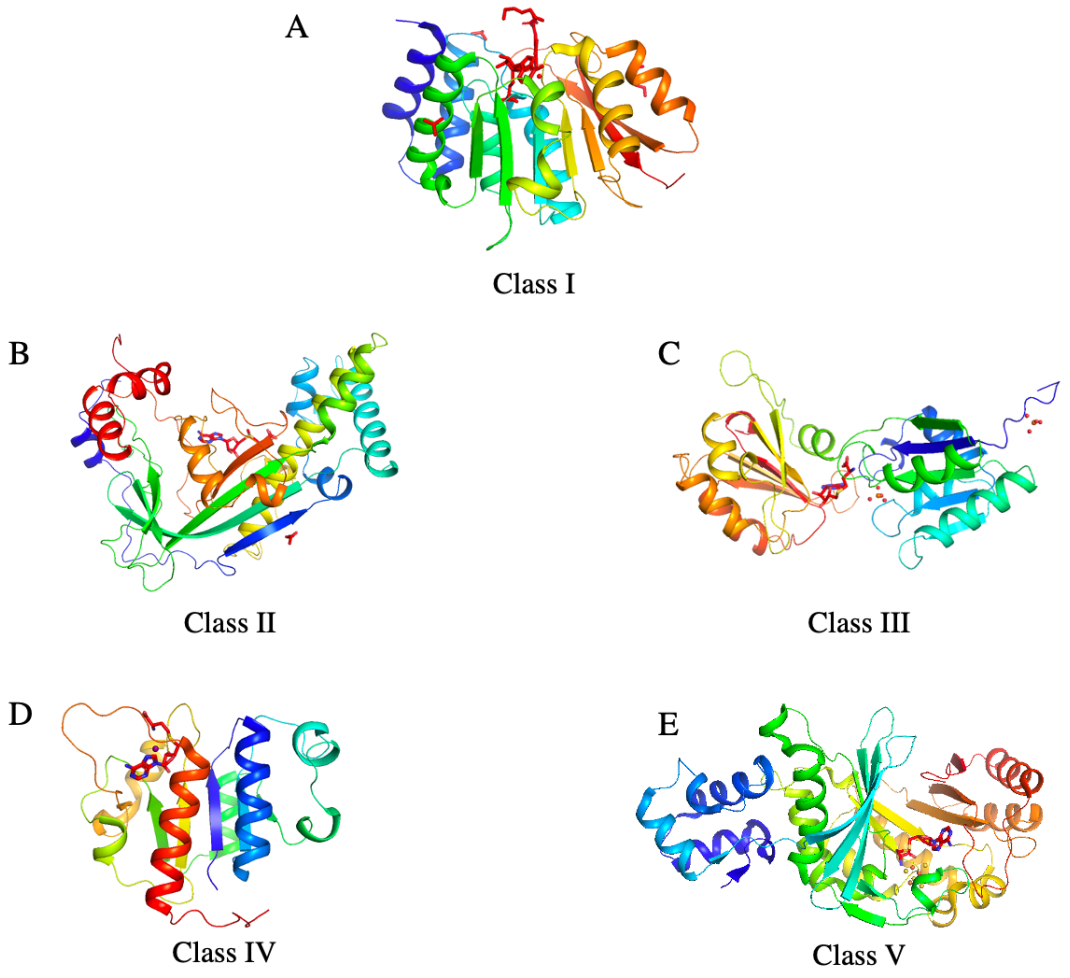


Figure 1.4 Representative structures of methyltransferases from class I to class V. (A) catechol-O-methyltransferase, PDB code: 6LFE. (B) methionine synthase, reactivation domain, PDB code: 1CBF. (C) cobaltprecorrin-4-methyltransferase, PDB code: 1CBF. (D) tRNA (cytidine (34)-2'-O)-methyltransferase, PDB code: 1MXI. (E) SET7/9, PDB code: 1O9S.

Class II methyltransferases are known for their function of reactivating oxidized cobalamin. The central feature of this class methyltransferase is a long antiparallel β-sheet with a sharp kink near one end (Figure 1.4 B). Helices surround the sheet above and below and extend beyond both ends, forming a horseshoe-like topology. The cofactor SAM always binds in an elongated conformation, in a groove by the inner edge of the central β-sheet.[41]

Class III methyltransferases primarily target tetrapyrroles, exemplified by cobalt-precorrin-4 MT, a key player in anaerobic vitamin B12 biosynthesis.[42] They are dimeric enzymes with active sites located between two similar but topologically distinct lobes (Figure 1.4 C). In solution, these proteins form dimers across the bridge between the domains, maintaining continuous β-sheets across the monomers.[43] SAM binds in a deep pocket near the N-terminal, adopting a tightly folded conformation that exposes the activated methyl group externally. While for the substrate binding region, the substrate is accommodated in a groove along the N-terminal domain.

Class IV methyltransferases are a part of the SPOUT enzyme superfamily, which derives its name from the bacterial tRNA methyltransferases SpoU (now known as TrmH) and TrmD.[44] These enzymes play a significant role in post-transcriptional modifications of transfer and ribosomal RNAs. Class IV methyltransferases are the second-largest group of methyltransferases, following class I. Their structural composition closely resembles that of class I, featuring a 6-strand parallel β-sheet surrounded by α-helices (Figure 1.4 D). Notably, five of these helices are situated on one side of the sheet, while the other two are on the opposite side. A distinctive "knot" topology, created by interactions between loops, serves as a defining characteristic of class IV methyltransferases.[45] SAM may stay in a bent conformation, nestled within a pocket formed by specific β-sheet strands and their associated loops.[46]

Class V methyltransferases are suppressors of variegation, enhancers of zeste and trithorax (SET) domains, that play a crucial role in methylating histone lysine and other proteins involved in transcriptional regulation across all life forms.[47] These enzymes have a relatively complex structure, featuring four α-helices within a sequence of twisted β-sheets (Figure 1.4 E). A pseudo-knot is formed by the loop preceding the C-terminal helix, distinct from class IV.[48] SAM always binds near this knot on the protein's surface. The histone binds on the opposite side, and the target lysine side chain approaches the SAM methyl through a narrow channel between the faces. Hydrogen bonding of the methionine amine to a conserved asparagine is thought to maintain the cofactor in a tightly folded conformation, with a tyrosine residue catalyzing the methyl transfer. The evolution of the SET domain superfamily appears to involve domain duplication in eukaryotes before lateral transfer to bacteria.[49]

There are many methyl donors in nature, such as vitamin B12, folate, and betaine.[50] Among them, SAM is the classical methyl donor for a great number of biochemical events.[51] The general

mechanism for methylation is an S_N2-like nucleophilic attack where the methionine sulfur acts as the leaving group, and the methyl group attached to it serves as the electrophile. During the methyl transfer step, the methyl acceptor in the substrate, the transferable methyl, and the sulfur atom in the donor SAM are presented in a linear position (around 3 Å to the methyl carbon and 4 Å between the acceptor and sulfonium group) as required. After this process, SAM converts to *S*-adenosyl-L-homocysteine (SAH). Hegazi et al. first reported that an *O*-methyltransferase works in S_N2-like transition in 1976.[52] Over the past few decades, other groups also have worked on the mechanism of methylation and found the reaction followed an S_N2-like transition.[53-55]

There is no doubt that methylation is important in extensive biological processes. Importantly, methylation also holds significance in organic synthesis. The process of methylating hydroxy and amino groups, thiols, and reactive carbons is a fundamental synthetic reaction. However, conventional synthetic techniques often rely on highly toxic reagents. These reactions, along with all nucleophilic substitution processes, are recognized as challenging when conducted on an industrial scale. In 2007, major pharmaceutical companies identified these reactions as a top priority for substitution with more environmentally friendly and catalytic approaches.[56] Although studies on *C*-methylation have increased over the past decade, some challenges still remain in organic synthesis, such as high catalyst loading, low regiospecific selectivity, and complex reaction conditions.[57] Therefore, effectively introducing a methyl group into aromatic systems has been a research hot spot in organic synthesis and drug discovery. *C*-methyltransferase is a potential approach to address these issues.

In the biosynthetic gene cluster of furaquinocin, Fur6 is a *C*-methyltransferase catalyzes the methylation of 1,2,4,5,7-pentahydroxynaphthalene (PHN) using SAM as a cofactor to yield 3-methyl-1,2,4,5,7-pentahydroxynaphthalene (3-Me-PHN) and SAH (Figure 1.5). The crystal structure of Fur6 with SAH was determined by Noguchi, T. et al.[34] The three-dimensional structure of Fur6 exhibits a classical Rossmann-like fold (Class I) comprising seven β strands (Figure 1.5), which is similar to other *C*-methyltransferases, such as CouO from *Streptomyces rishiriensis*[58] and BezA from *Streptomyces* sp. RI18 (Figure 1.6 A, B).[59] However, when the active site of Fur6 is compared to that of CouO, the crucial residues His120 and Arg121 found in CouO which interacted with substrate through hydrogen bonds are not conserved in Fur6 (Figure 1.6 C). Furthermore, the residue Glu170 which functions as the catalytic base in the reaction catalyzed by BezA is also not conserved in Fur6 (Figure 1.6 D). These differences between Fur6 and its homologous proteins make it difficult to infer the catalytic mechanism of Fur6. Moreover, due to the O_2 lability of the substrate PHN, the 3-Me-PHN cannot be detected, which leads to the catalytic mechanism of Fur6 being difficult to clarify through experimental methods.

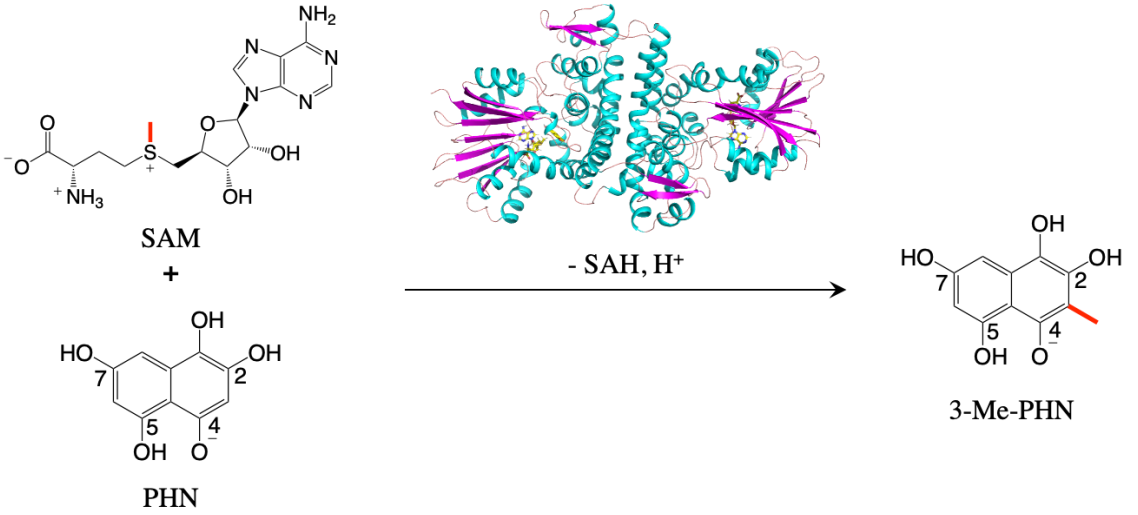


Figure 1.5 Schematic representation of methylation catalyzed by Fur6 complexed with SAM and PHN.

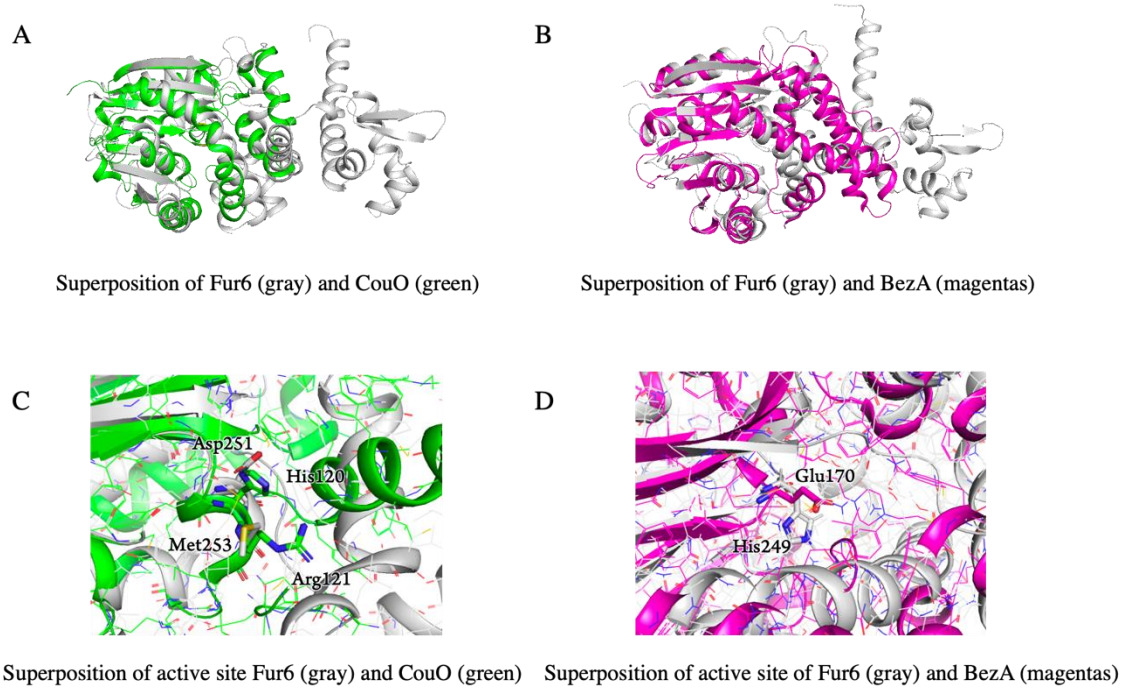


Figure 1.6 Superpositions of Fur6 and its homologues. All the proteins are shown in monomer. (A) Superposition of Fur6 (gray) and CouO (green). (B) Superposition of Fur6 (gray) and BezA (magentas). (C) Superposition of active site Fur6 and CouO. (D) Superposition of active site of Fur6and BezA.

1.1.3 Prenyltransferases are Key in Metabolic Pathways and Structural Diversity

Prenyltransferase is a crucial enzyme involved in the modification of small molecules and proteins. This class of enzymes facilitates the transfer of the prenyl groups (such as farnesyl or geranylgeranyl) to specific target molecules or proteins, a process known as prenylation, and the molecules with a 5-carbon geranyl unit belong to the class of terpenoids. Prenylation brings diversity to natural products in structure and bioactivity.

All the terpenoid compounds are derived from linear prenyl diphosphate molecules, the building blocks are isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) and are produced from mevalonate (MVA) and methylerythritol phosphate (MEP) pathways (Figure 1.7). The MVA pathway is present in animals, fungi, archaea, and some bacteria, while the MEP

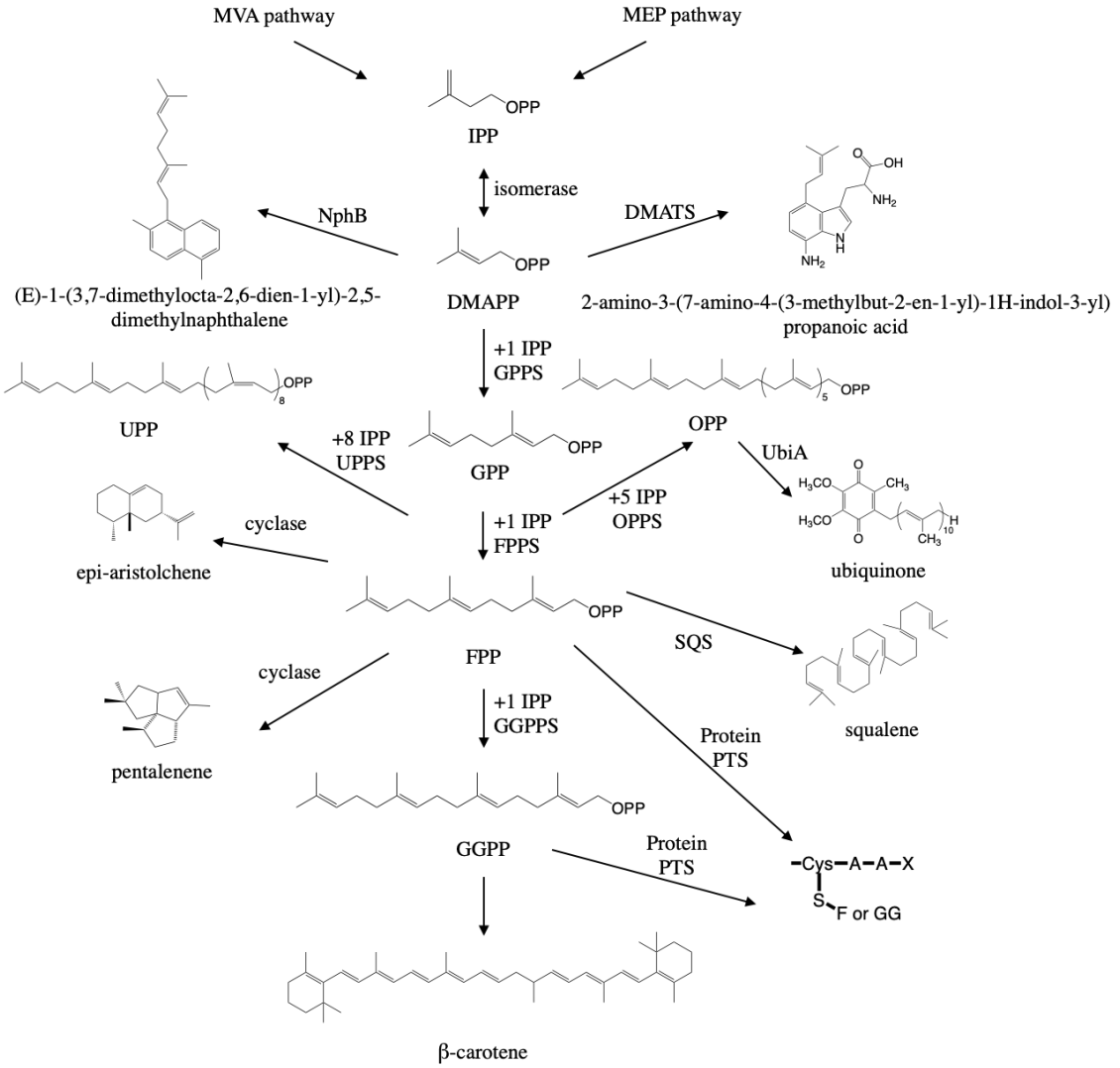


Figure 1.7 Prenyltransferases have broadly effect in various natural product biosynthesis.

pathway is the primary route for most bacteria to synthesize isoprene.[60] This pathway, discovered in the 1950s, is used for producing isoprene units in liver tissues and yeast.[61] Prenyltransferases play a vital role in the condensation of IPP and DMAPP to create various isoprenoid diphosphates with different carbon chain lengths. This includes 10-carbon geranyl diphosphate (GPP), 15-carbon farnesyl diphosphate (FPP), and 20-carbon geranylgeranyl diphosphate (GGPP) (Figure 1.7).

Prenyltransferases can be divided into six classes depending on their catalytic mechanism.[62] Class I prenyltransferase, named head-to-tail prenyltransferase, catalyzes the condensation of linear isoprenyl diphosphates through a consecutive head-to-tail process.[63] Class II prenyltransferase, known as head-to-head prenyltransferase, mainly catalyzes the condensation of two FPP molecules in a head-to-head fashion to form presqualene diphosphate. Class III prenyltransferase, named head-to-middle prenyltransferase, catalyzes the head-to-middle condensation reaction between the C2-C3 double bond of the first allylic substrate and C1 in a second allylic substrate, subsequently producing branched mono- or sesquiterpenes. Class IV prenyltransferase, called terpenoid cyclase, catalyzes the cyclization of linear isoprenyl pyrophosphates, such as GPP, FPP, and GGPP, into various cyclic isoprenoid compounds. Class V prenyltransferase, an aromatic prenyltransferase, facilitates the transfer of isoprenoid diphosphates to aromatic compounds, a reaction corresponding to Friedel–Crafts alkylation.[64] The last class of prenyltransferase catalyzes protein prenylation, a post-translational modification of proteins that plays a fundamental role in regulating cellular functions like signal transduction and intracellular trafficking[62] (Figure 1.7).

In primary and secondary metabolites of fungi, bacteria, and plants, aromatic prenyltransferases are pivotal in the formation of primary and secondary metabolites in fungi, bacteria, and plants. The reaction catalyzed by aromatic prenyltransferases always follows an S_N2-like mechanism, nucleophilic attack on the carbon atom of prenyl donor substrate, and the diphosphate moiety serves as a leaving group. This superfamily is typically categorized into three groups, distinguished by their secondary metabolisms and biological activity localization: the cytosolic ABBA (α-β-β-α barrel) group, the dimethylallyl tryptophan synthase (DMATS) group, the membrane-embedded UbiA type.[65] NphB, previously referred to as Orf2, was initially identified as the first ABBA-type prenyltransferase in the biosynthesis of the antioxidant naphterpin within *Streptomyces* sp. (strain CL190) by Kuzuyama, T. et al (Figure 1.8 A).[30] NphB exhibits

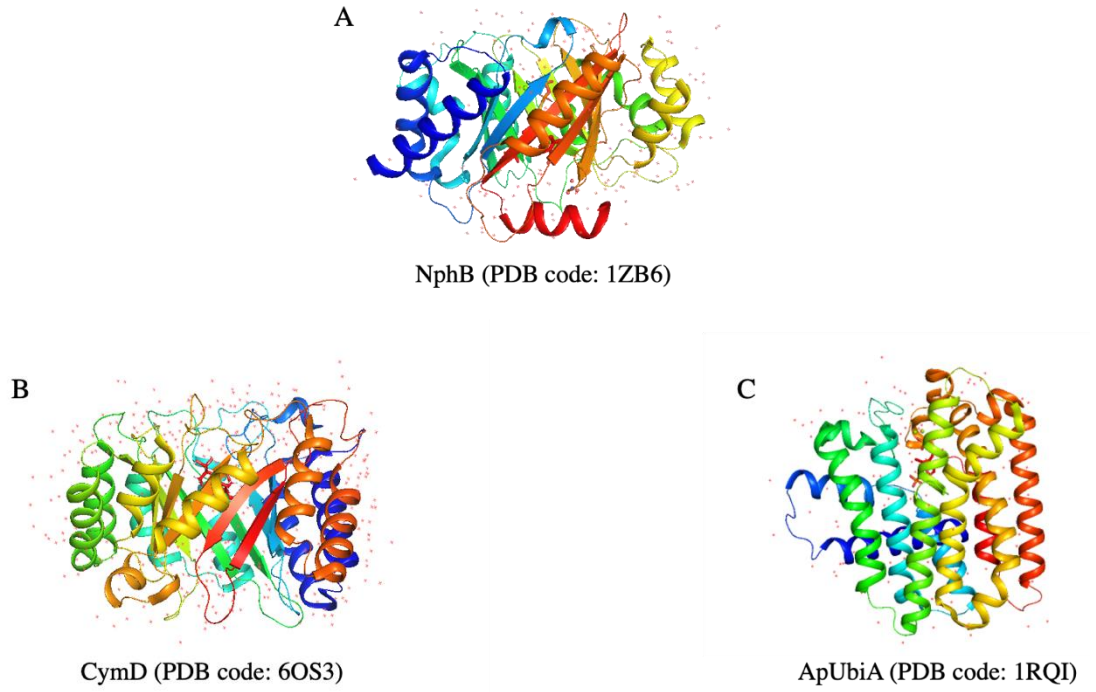


Figure 1.8 Representative structures of aromatic prenyltransferases. (A) cytosolic ABBA (α-β-β-α barrel) prenyltransferase, NphB, PDB code: 1ZB6. (B) DMATS prenyltransferase, CymD, PDB code: 6OS3. (C) UbiA prenyltransferase, ApUbiA, PDB code: 1RQI.

remarkable substrate specificity to aromatic substrates, which interact with GPP to yield a variety of *O*- or *C*-prenylated aromatic compounds. The three-dimensional crystal structure of NphB reveals a 10-stranded antiparallel β-barrel fold comprising five repetitive αββα elements, with the catalytic site situated in the center of the β-barrel. NphB uses a magnesium ion (Mg^{2+}) to catalyze isoprenyl diphosphate substrates. In contrast, ABBA-type enzymes like CloQ and DMATS do not rely on Mg^{2+} for prenylation. CloQ, isolated from *Streptomyces roseochromogenes*, plays a role in clorobiocin biosynthesis, an aminocoumarin antibacterial agent that inhibits DNA gyrase. A

primary and tertiary structural analysis conducted by Metzger, U. et al. has revealed that CloQ lacks the aspartate-rich metal binding motif for metal ion binding and remains active in the absence of divalent cations,[66] suggesting a substrate ionization mechanism based on hydrogen bond interaction. DMATS plays a crucial role in ergot alkaloid biosynthesis and produces secondary metabolites with significant importance in toxicology and pharmacology. It was initially isolated from *Claviceps purpurea* by Tudzynski, P. et al (Figure 1.8 B).[67] Despite showing no amino acid sequence homology with NphB, CloQ, or their orthologs, DMATS shares a β-barrel fold conformation with the ABBA-type superfamily according to the crystal structure studies.[66] Unlike cytosolic ABBA prenyltransferases, UbiA subfamily enzymes are membrane proteins, catalyzing essential steps in lipoquinone production. The crystal structure of ApUBiA (Figure 1.8 C) shows the protein possesses two aspartate-rich motifs in complexes with GPP or DMAPP and Mg^{2+}, similar to other prenyltransferases that have metal-binding motifs.[68]

Although I have introduced several research on prenyltransferases above, in fact, over 80,000 terpenoid compounds have been characterized, but only a few prenyltransferases are known.[62] Further investigations are needed to uncover new enzyme functions, structures, and the gene clusters responsible for terpenoid biosynthesis. Notably, the therapeutic potential of terpenoids in treating human diseases like cancer, hepatitis, and osteoporosis has garnered increased interest.[7,69] Exploring medicines with fewer side effects from natural products holds promise for developing a novel class of low-toxicity drugs.

In the family of aromatic prenyltransferases, Fur7 is a member of the furaquinocin biosynthetic gene cluster, which was introduced earlier. The reaction catalyzed by Fur7 was demonstrated by Kumano, T. et al. in 2010.[20] Fur7 transfers a 10-carbon geranyl group using GPP as the prenyl donor to the polyketide skeleton 2-methoxy-3-methyl-flaviolin (MMF). This leads to the production of 6-prenyl-2-methoxy-3-methyl-flaviolin (C6-product) and 7-*O*-geranyl-2-methoxy-3-methyl-flaviolin (O7-product) in a 10:1 ratio (Figure 1.9). Notably, unlike its homologs NphB

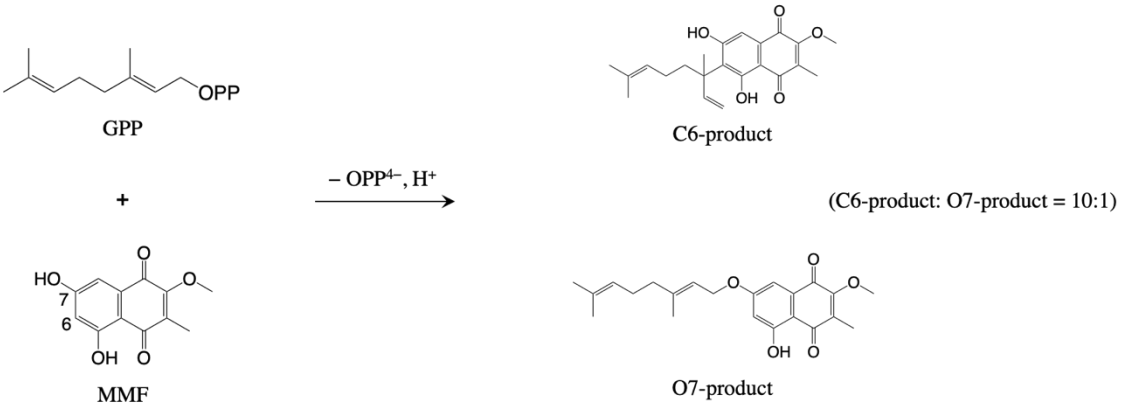


Figure 1.9 Reaction catalyzed by Fur7. (Reported by Kumano, T. et al.)

or CloQ, Fur7 predominantly catalyzes a reaction at C6 of the substrate to generate the "reverse" product. The products produced by the reaction that occurs at C1 of the prenyl group are termed the normal product, whereas those produced by the reaction that occurs at C3 of the prenyl group are called the reverse product. Moreover, Fur7 displays low specificity for both prenyl donor and prenyl acceptor substrates. In contrast to its homolog NphB, which requires Mg^{2+} for catalysis, Fur7 can function without the divalent cations. The addition of 5 mM $MgCl_2 \bullet 6H_2O$ only resulted in a 30% increase in activity.[20] However, Noguchi, T. et al. recently reported different results compared to the findings of Kumano, T. et al. They proposed that the authentic substrate of Fur7 is 6-methoxy-7-methylnaphthalene-1,3,5,8-tetraol (MMT), and its product is 6-prenyl-2-methoxy-3-methyl PHN, which is subsequently oxidized to C6-product (Figure 1.10). Notably, the activity of Fur7 is significantly enhanced under reducing conditions, and the O7-product was not detected during their experiments.[34] However, they could not observe the reduced C6-product and the crystal structure remains unsolved.

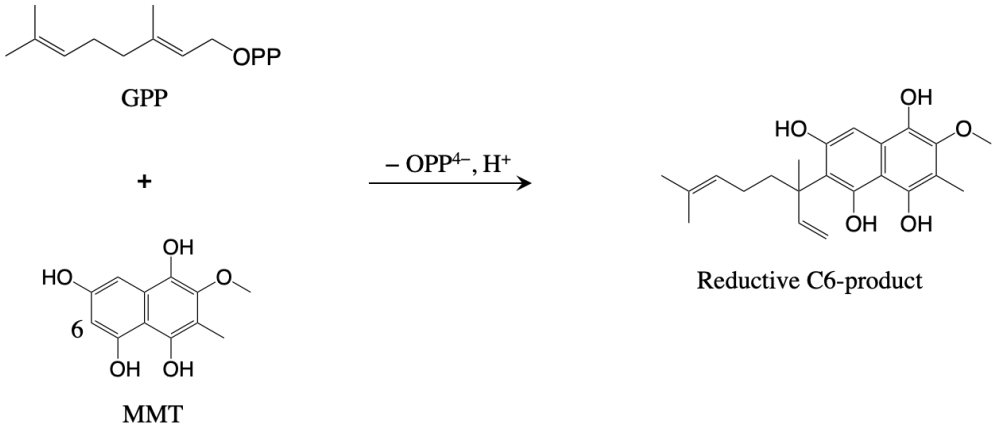


Figure 1.10 Reaction catalyzed by Fur7. (Reported by Noguchi, T. et al.)

1.1.4 Computational Tools Demonstrate Their Extensive Utility in Enzymatic Study

In recent decades, computational tools have shown potential in enzymatic study, such as mechanistic study, structure modeling, and rational design of enzymes. They sometimes could even go beyond the limitations of experimental methodologies.[70-73] Furthermore, computational methods offer the advantage of significantly reducing both time and experimental costs. These tools provide the means to conduct in-depth atomic-level analysis of enzymes. For example, they allow researchers to investigate how specific residues can alter enzymatic chemo-, regio-, or stereo-selectivities. They also help researchers understand the effects of amino acid substitutions on protein conformational dynamics and how these changes impact substrate binding. There is no doubt that computational techniques empower researchers to explore enzyme behavior with precision and efficiency.

Quantum Mechanics (QM) calculations encompass a diverse array of approximations, all aimed at tackling the Schrödinger equation to gain insights into the molecular properties and energetics of a specific chemical system (Figure 1.11). QM calculations consist of all the *ab initio* methods based on the Hartree–Fock (HF) approximation, semiempirical approaches, and the widely adopted Density Functional Theory (DFT).[74] DFT offers remarkable precision in

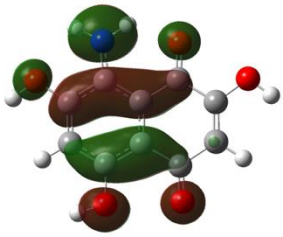


Figure 1.11 Example of QM calculations: the image of HOMO calculation of 8-AF. Oxygen, nitrogen, and hydrogen atoms are shown in red, blue, and white, respectively. Molecular orbital is shown in shadow.

characterizing chemical systems. Sato, H. et al used the DFT calculations to clarify the mechanism of skeletal rearrangement reaction involved in scalarane-type sesterterpenoid biosynthesis,[75] Pulido, A. et al. combined experimental and DFT study elucidated the ketonic decarboxylation reaction mechanism,[76] and Fomina, L. et al conducted the DFT study on the reaction mechanism of Glaser oxidative coupling of terminal acetylenes.[77] Ohashi, M. et al performed DFT study and *in vitro* enzymatic characterization to elucidate the mechanistic insight of *O*-methyltransferase-like protein LepI.[78] Levandowski, B. J. et al used DFT method to demonstrate the hyperconjugative effects of Diels–Alder reactions of 5-substituted cyclopentadienes.[79] However, DFT calculation evidently comes with an exponentially escalating computational cost as the system size increases. This constraint results in the high cost associated with treating an entire protein using DFT, which accounts for the selective focus on a limited number of active site residues in QM-based methodologies.

The integration of quantum mechanics/molecular mechanics (QM/MM) techniques, which was initially pioneered by Warshel and Levitt,[80] has become a widely adopted approach for investigating complex biomolecular systems and drug exploration.[81] The profound impact and significance of this computational methodology were officially acknowledged in 2013 by the Nobel Prize in Chemistry. QM/MM methods were initially conceived to facilitate the exploration of chemical reactions necessitating quantum mechanical treatment, such as bond cleavage or formation reactions, which are too extensive for in-depth analysis at a high level of theory. Consequently, QM/MM calculations conduct the precise treatment of a small region of the chemical system, for example, the active site of an enzyme by using QM, while the remaining components are described using computationally efficient molecular mechanics (MM) methods (Figure 1.12). MM methods employ classical mechanics to simulate atom–atom interactions,

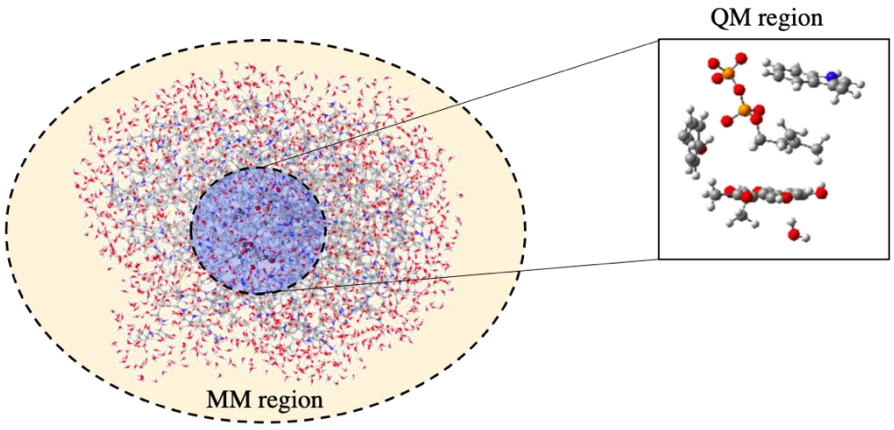


Figure 1.12 Concept map of QM/MM calculation. The QM and MM region are circled by blue and yellow, respectively.

calculating the energy of the system through potential energy functions referred to as force fields.[82] In the study on biocatalysis, QM/MM methods offer the capability to investigate chemical reactions involving millions of atoms at atomic resolution. This approach considers the intricate effects of the solvent environment and the protein molecule, demonstrating their influence on the reaction pathway. Dong, L. and Liu, Y. use the QM/MM strategy to elucidate the catalytic mechanism of an aminodeoxychorismate lyase, PabC which catalyzes a key step for the biosynthesis of folate.[83] Moriwaki, Y. et al. use the QM/MM methods to illustrate the molecular mechanism of *p*-hydroxybenzoate hydroxylase mutants which possess elevated activity in producing gallic acid.[71] Viegas, M. et al. perform QM/MM methods to demonstrate the catalytic mechanism of thermophilic glucuronoyl esterase, which works for the biomass treatment.[84]

Quynh N, N et al conduct a QM/MM study to reveal that polyanionic gallium-based cages accelerate cyclization reactions of pentadienyl alcohols.[85] Obviously, over the past decade, more and more researchers have used the QM/MM methods in their studies.

The analysis of enzyme conformational flexibility is a critical component in structural studies, and this can be achieved through a sampling method called molecular dynamics (MD) simulations (Figure 1.13). In MD simulations, ensembles of conformations are generated through integrating

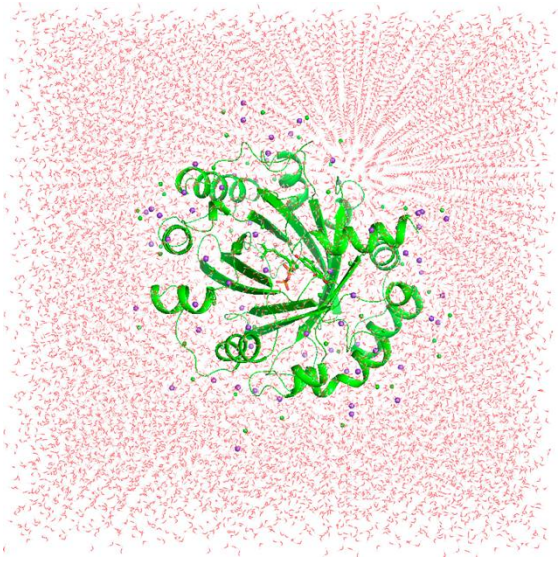


Figure 1.13 Concept map of classical MD simulations. The protein is show in green, Na^+ and Cl^- are shown in purple and green ball, respectively. Water molecules are shown in stick model, oxygen and hydrogen atoms are shown in red and white, respectively.

equations of motion. MD simulations can be classified into quantum or classical simulations based on the nature of the Hamiltonian they employed. *Ab initio* MD simulations is an important theoretical tool that generates finite-temperature dynamical trajectories through forces obtained directly from electronic structure calculations conducted during simulations. *Ab initio* MD simulations allows for studying chemical bond-breaking and forming processes and accounts for electronic polarization effects. It is widely applied in physics, chemistry, and biology. Classical MD simulations, conversely, are the predominant strategy for investigating protein conformational flexibility. In this method, both the protein and the solvent are represented using force fields described by simple potentials. Integration of these potentials yields the forces acting on each atom, enabling the computation of corresponding accelerations, and subsequently, the determination of new velocities and positions. This iterative process produces a trajectory of the protein, offering insights into enzyme conformational dynamics in relation to simulation time.[74] MD simulations provide the opportunity for researchers to analyze the biochemical processes

such as ligand binding, protein folding, voltage- or ligand-induced conformational change, or membrane transport.

As previously mentioned, numerous computational strategies can be applied in biochemical studies to assist researchers in unraveling the mysteries of enzymes. In the enzymatic study, alongside these strategies, another challenging problem is protein structure prediction, which has perplexed computational biology and chemistry for over half a century.[86] In December 2020, AlphaFold2 won the championship in the 14th Critical Assessment of Structure Prediction.[87] AlphaFold2 is a groundbreaking artificial intelligence system developed by DeepMind, which stands out for its remarkable ability to accurately predict 3D protein structures from amino acid sequences at the atomic level.[88] This groundbreaking achievement has made a significant impact on the field of protein structure prediction and has garnered substantial attention. Furthermore, the subsequent release of 3D structures for over 200 million proteins, forecasted by AlphaFold2,[89] has generated great excitement in the scientific community, especially in the fields of biology and medicine. AlphaFold2 shows a significant influence on structural biology, as well as the research related to protein structures, such as drug discovery, protein engineering, and protein function prediction. Hu, L. et al. combine X-ray crystallography and AlphaFold2 prediction to determine the structure of the VP8* domain (VP8*B) of VP4, which is a spike protein and belongs to group B rotaviruses.[90] Hutin, S. et al. clarified the structure of the vaccinia virus DNA helicase through the cryo-EM method and AlphaFold2 prediction.[91] Zhang, Y. et al. utilize a molecular docking program Glide and AlphaFold2 to get the comparable results as using the experimental structures for drug virtual screening.[92] Although AlphaFold2 was only released three years ago, it has already been applied in extensive research fields.

The computational strategies introduced above have their own strengths and weaknesses. It is conceivable that by judiciously combining these methods with experimental tools, researchers can address the challenge more efficiently.

1.2 Objectives

In this research, I employ computational analysis to investigate the chemistry of *C*-methyltransferase Fur6 and aromatic prenyltransferase Fur7. My objectives for the study on Fur6 include elucidating the binding mode of the Fur6-PHN-SAM complex, understanding the enzymatic mechanism, and exploring the substrate specificity of Fur6. For the study on Fur7, my aim is to uncover the original substrate of Fur7, determine the model of its 3D structure, analyze the binding mode of the Fur7-MMT-GPP complex, and map out the reaction pathway.

I anticipate that my comprehensive investigation will provide valuable insights into the Fur6 and Fur7 chemistry, deepening the understanding of their molecular mechanisms, and expanding

the scope of research on *C*-methyltransferase and aromatic prenyltransferase. Furthermore, this research offers an opportunity to engineer these enzymes, potentially leading to enhanced yields of the desired products. This could have significant implications for various applications, such as improving the low-toxicity processes in pharmaceutical synthesis and optimizing synthetic routes more efficiently.

2 Theory Framework

2.1 MD Simulations

2.1.1 Theory of MD Simulations

The method employed for simulating atoms and molecules on the computer is known as MD simulations. During a specific timeframe, atoms and molecules interact, offering insights into the dynamic "evolution" of the system. In the most widely used version, Newton's equations of motion for a system of interacting particles are numerically solved to determine the trajectories of atoms and molecules. MD simulations numerically solve the motion equations for each atom or molecule constituting the material system, tracking changes in position, velocity, and energy over time. The forces between particles and their potential energies are typically calculated using molecular mechanical force fields or interatomic potentials. MD simulations prove beneficial when examining temporal changes and non-steady states, making it essential for observing structural alterations. Chemical physics, materials science, and biophysics are the primary fields in which this method finds application.

2.1.2 Equations of Motion

To integrate the equation of motion, firstly, some basic variables need to be set. A molecule is composed of various atoms. At a certain time t, a given molecular geometry (R) composed of N atoms can be expressed as a function of the N atomic positions (r_N).

$$R = (r_1, r_2, \dots, r_N) \quad (2.1)$$

Successive configurations of the system are obtained by integrating Newton's laws of motion for each atom.

$$(F = ma) \quad (2.2)$$

The acceleration a is the second derivative of the position r with respect to time t. Therefore, each atom can be described as the equation below. F_i means the force acting on the i-th atom.

$$F_i = m_i \frac{\partial^2 r_i}{\partial t^2} \quad (2.3)$$

While the force also could be represented as the opposite of the derivative of the potential energy V.

$$F_i = -\frac{\partial V}{\partial r_i} \quad (2.4)$$

To integrate the equations above, a numerical approach named finite difference method is applied. The main idea is that the integration can be divided into many small finite time steps δt. The time step needs to be small to describe the process of interest accurately. At each step, all the N equations are simultaneously solved and the forces acting on each atom $(F_1, \dots, F_N)$ at time t are determined. Depending on the force, the acceleration could be calculated.

$$a_i = \frac{F_i}{m_i} = -\frac{1}{m_i}\frac{\partial V}{\partial r_i} \quad (2.5)$$

Given the molecular position, velocity, and acceleration at time t, the position and velocity at the following time step $(t + \delta t)$ can be predicted. This process is employed across multiple steps (Figure 2.1), resulting in a trajectory that delineates the variations in both the positions and velocities of a particle over time.

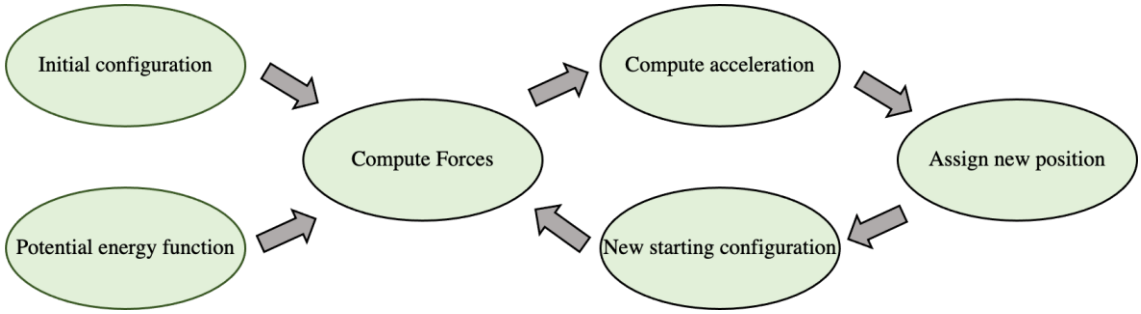


Figure 2.1 Scheme of MD simulation algorithm.

To integrate the equations of motion, the position and the dynamical properties of a molecule at time $t + \delta t$ are approximated through a Taylor expansion at time t.

$$\begin{aligned} r(t+\delta t) &= r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \cdots \\ v(t+\delta t) &= v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \cdots \\ a(t+\delta t) &= a(t) + \delta t b(t) + \cdots \end{aligned} \quad (2.6)$$

In equation (2.7), v is the velocity, a is the acceleration, b is the third derivative of the position with respect to time. Verlet algorithm is the most widely used numerical method. In this method, the current positions $r(t)$ and accelerations $a(t)$, and the positions from the previous step $r(t - \delta t)$ to retrieve the positions in the following step $r(t + \delta t)$. The algorithm creates by simply writing the equations for the past and future positions:

$$\begin{aligned} \mathrm{r}(t+\delta t) &= \mathrm{r}(t) + \delta t \mathrm{v}(t) + \frac{1}{2}\delta t^2 \mathrm{a}(t) + \cdots \\ \mathrm{r}(t-\delta t) &= \mathrm{r}(t) - \delta t \mathrm{v}(t) + \frac{1}{2}\delta t^2 \mathrm{a}(t) + \cdots \end{aligned} \quad (2.8)$$

Subsequently, further adding these two terms above the basic Verlet algorithm equation could be obtained.

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) \quad (2.9)$$

The basic Verlet algorithm equation has been further developed, the velocity Verlet method is built, which becomes the most widely used method in MD simulation. It derives from the position and the velocity Taylor expansions:

$$\begin{aligned} r(t+\delta t) &= r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) \\ v(t+\delta t) &= v(t) + \frac{1}{2}\delta t[a(t) + a(t+\delta t)] \end{aligned} \quad (2.10)$$

by using the velocity Verlet method, the motion of the molecule through the current position $r(t)$, velocity $v(t)$, and acceleration $a(t)$ can be calculated (Figure 2.2).

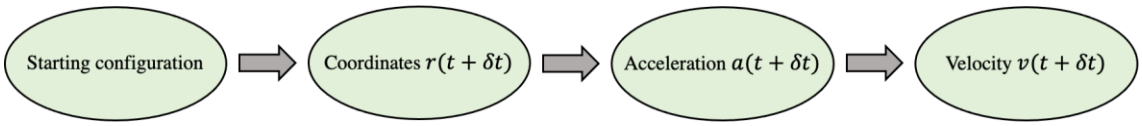


Figure 2.2 Scheme of Velocity Verlet.

2.1.3 Molecular Force Fields

In the realm of chemistry and molecular modeling, a force field is a computational technique employed to estimate the intermolecular forces and intramolecular forces. Specifically, the term "force field" encompasses the functional structure and parameter sets utilized to compute the potential energy of a system comprising atoms or coarse-grained particles in molecular mechanics, molecular dynamics. The parameters for a selected energy function can be derived from experimental data in physics and chemistry, quantum mechanics calculations, or a combination of both.

The force field can be typically expressed as:

$$\begin{aligned} U = &\sum_{\text{bonds}} \frac{1}{2}k_b(r-r_0)^2 + \sum_{\text{angles}} \frac{1}{2}k_a(\theta-\theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2}[1+\cos(n\phi-\delta)] \\ &+ \sum_{\text{improper}} V_{imp} + \sum_{\text{LJ}} 4\epsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}}, \end{aligned} \quad (2.11)$$

the initial four terms account for intramolecular or local contributions to the total energy, encompassing bond stretching, angle bending, dihedral, and improper torsions. The last two terms, characterize the repulsive and van der Waals interactions, along with the Coulombic interactions.

The first term is bond stretching, which is frequently represented using a simple harmonic function governing the length of covalent bonds. Appropriate values for r_0 can be derived from X-ray diffraction experiments.

The second term is angle bending, which is also usually represented by a harmonic potential as shown below.

$$U_{\text{bending}} = \frac{1}{2} k_a (\cos\theta - \cos\theta_0)^2 \qquad (2.12)$$

The third term is torsional energy, which is usually represented by a cosine function, ϕ is the torsional angle, δ is the phase, n the number of minima or maxima between 0 and 2 π, and V_n is the height of the potential barrier.

The fourth term is improper torsions, which can be expressed as the equation as followed. The ω is the improper angle corresponding to the deviation from planarity.

$$U_{\text{imp}} = \sum_{\text{impropers}} \frac{k_{\text{imp}}}{2} [1 + \cos(2\omega - \pi)] \qquad (2.13)$$

The fifth term represents Van der Waals interactions, these interactions between two atoms originate from the equilibrium between repulsive and attractive forces. The repulsion is a consequence of the electron clouds of both atoms overlapping, whereas the attractive component arises from interactions between induced dipoles and can be represented as r^{-6}. The 12-6 Lennard-Jones potential, noted as LJ, is very often used to represent these interactions.

The last term in the equation (2.10) represents the electrostatic interactions. The common approach for calculating partial atomic charges in molecular simulations involves assigning charges to each nucleus based on Coulomb's law. While fitting to experimental data is practical for small molecules, the most reliable method is to perform an *ab initio* calculation and derive charges from the quantum mechanical potential.

2.2 QM/MM Simulations

2.2.1 Theory of QM/MM Simulations

As I already briefly introduced in Section 1.1.4, in the study of biochemical systems like enzymes, full quantum mechanical descriptions are often impractical due to their size, while

molecular mechanics force fields lack flexibility for processes involving bond formation or breaking. To address this, a hybrid approach, known as QM/MM simulations, has been developed (Figure 1.12). This method combines quantum chemistry for a small part of the system with computationally more efficient molecular mechanics for the larger part. This allows for a more realistic representation of complex processes in large biochemical systems.

The hybrid QM/MM potential energy involves three types of interactions: within the QM region, within the MM region, and between QM and MM atoms. Describing interactions within QM and MM regions is straightforward at their respective levels. However, interactions between the two subsystems pose challenges, leading to various proposed approaches. These can be broadly categorized as additive and subtractive coupling schemes.

2.2.2 Additive QM/MM Coupling Scheme

In additive schemes, the larger MM system encompasses the QM system, and the total potential energy of the entire system is the sum of MM energy terms, QM energy terms, and QM/MM coupling terms.

$$V_{\mathrm{QM/MM}} = V_{\mathrm{QM}}(\mathrm{QM}) + V_{\mathrm{MM}}(\mathrm{MM}) + V_{\mathrm{QM-MM}}(\mathrm{QM} + \mathrm{MM}) \qquad (2.14)$$

The interactions between the QM and MM region can be described in various approaches. Such as mechanical embedding, electrostatic embedding, and polarization embedding. Among them, mechanical embedding is the most basic approach, in which all interactions between the two subsystems are handled at the force field level. Electrostatic embedding is an enhancement to mechanical embedding that involves incorporating polarization effects. In the electrostatic embedding scheme, the computation of the electronic wave function addresses the electrostatic interactions between the two subsystems. The electrostatic embedding scheme incorporates the partial charges of the MM region into the quantum mechanical Hamiltonian, providing a better description of the electrostatic interaction between the QM and MM regions and allows the QM wavefunction to be polarized. Polarization embedding is a higher level of sophistication that involves incorporating the polarizability of MM atoms. In the polarization embedding scheme, both regions can reciprocally induce polarization. Consequently, not only is the QM region polarized by MM atoms, but the QM region can also prompt polarization in the MM system. Various approaches have been devised to simulate the polarization of MM atoms. Some widely employed methods include the charge-on-a-spring model, the induced dipole model, and the fluctuating charge model.[93]

2.2.3 Subtractive QM/MM Coupling Scheme

In the subtractive approach, the QM/MM energy of the system is determined through a three-step process. Initially, the energy of the entire system, encompassing both QM and MM regions, is assessed at the MM level. Subsequently, the QM energy of the isolated QM subsystem is incorporated. In the third step, the MM energy of the QM subsystem is calculated and then subtracted. This final step serves to rectify the double counting of interactions within the QM subsystem (Figure 2.3).

$$V_{\mathrm{QM/MM}} = V_{\mathrm{MM}}(\mathrm{MM}+\mathrm{QM}) + V_{\mathrm{QM}}(\mathrm{QM}) - V_{\mathrm{MM}}(\mathrm{QM}) \qquad (2.15)$$

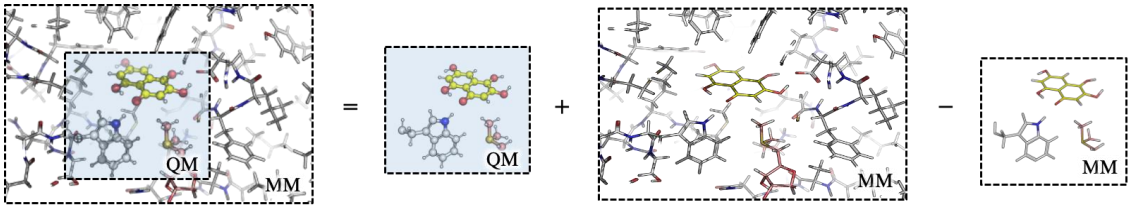


Figure 2.3 Subtractive QM/MM coupling. The atoms are calculated under MM and QM levels are shown in lines and ball-stick mode, respectively.

The abbreviations QM and MM refer to the atoms within the QM and MM subsystems, respectively. The subscripts denote the level of theory used for computing potential energies V. The ONIOM method, pioneered by the Morokuma group,[94] is the prevailing subtractive QM/MM scheme in use. The subtractive QM/MM coupling scheme offers a primary advantage in that it eliminates the need for communication between the quantum chemistry and molecular mechanics routines. This simplifies the implementation process significantly.

2.2.4 Capping Bonds at the QM/MM Boundary

When the QM and MM subsystems are linked by chemical bonds, caution must be exercised during the evaluation of the QM wave function. A direct severance of the QM/MM bond can result in one or more unpaired electrons in the QM subsystem. In actuality, these electrons form pairs in bonding orbitals with electrons from the atom on the MM side. To address the impact of these open valences, the simplest solution is to introduce a monovalent link atom at a suitable position along the bond vector between the QM and MM atoms.

Although hydrogen is commonly employed, there are no restrictions on the type of link atom, and complete fragments, such as methyl groups, can also serve as caps for the QM subsystem. These link atoms exclusively exist in the QM calculation and remain invisible to the molecular MM atoms. In theory, each link atom introduces three additional degrees of freedom to the system. However, in practice, the link atom is consistently positioned along the bond in every simulation

step, effectively eliminating these extra degrees of freedom. At each step, the force acting on the link atoms is proportionally distributed between the QM and MM atoms of the bond using the lever rule.

An alternative frequently used instead of the link atom scheme involves substituting a chemical bond between the QM and MM subsystem with a double-occupied molecular orbital. Originating from the pioneering research of Warshel and Levitt,[80] this concept assumes that the electronic structure of the bond remains unaffected by alterations in the QM region. The two predominant approaches are the localized hybrid orbital method, introducing orbitals at the QM atom, and the generalized hybrid orbital approach, incorporating additional orbitals on the MM atom.

2.3 AlphaFold2

2.3.1 Principle of AlphaFold2

As detailed in Section 1.1.4, AlphaFold2, developed by DeepMind, is a machine-learning-based model designed to predict protein structures. This innovative tool can accurately predict the 3D structures of proteins from amino acid sequences at the atomic level. Its foundation lies in cutting-edge deep learning algorithms and the conservation of protein structures throughout evolution.

AlphaFold2 applies recently developed deep learning algorithms, with a significant contribution from the attention mechanism-based transformer, enhancing its overall performance. The transformer is a novel deep neural network, that utilizes self-attention mechanisms to extract intrinsic features, showing considerable potential for diverse applications in the fields of artificial intelligence.

The biological principle utilized by AlphaFold2 is the conservation of protein structure in evolution. Proteins tend to be conservative in their evolution, with most mutations showing less impact on the function of the protein. Notably, protein structure exhibits greater conservatism than its amino acid sequence. The conservation in alignment often indicates its importance in protein folding or function. The co-evolution between two amino acid residues signifies their interaction, forming the basis for 3D structure prediction in AlphaFold2.

2.3.2 Architecture of AlphaFold2

AlphaFold2 incorporates an architecture that comprises three modules: the input module, the Evoformer module, and the structure module (Figure 2.4).

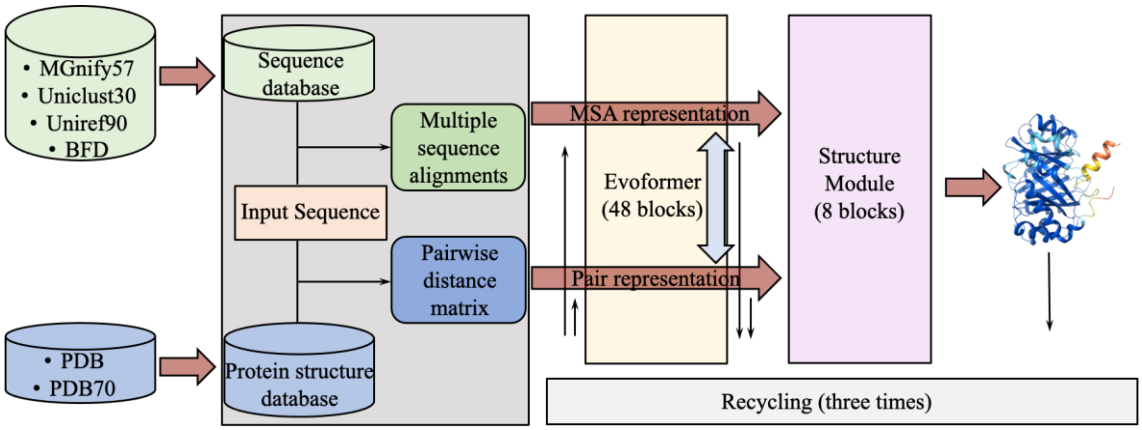


Figure 2.4 The architecture of AlphaFold2.

- The Input Module

In the input module, when provided with an amino acid sequence to AlphaFold2, it identifies homologs from sequence databases and conducts multiple sequence alignment (MSA) by aligning the input sequence with its homologs. The system also checks if any of the homologs possess a 3D structure in protein structure databases, then generates a pairwise distance matrix between amino acids. Subsequently, AlphaFold2 produces MSA representation and pair representation. In contrast to homology modeling, although both homology modeling and AlphaFold2 incorporate MSA, AlphaFold2 applies the co-evolution information derived from the MSA, while homology modeling does not utilize it. Moreover, AlphaFold2 employs several high-quality protein sequence databases, such as MGnify57, Uniclust30, Uniref90, and BFD (a database constructed by the AlphaFold team). For the structure databases that are used in training and as templates, AlphaFold2 adopts PDB and PDB70. AlphaFold2 also incorporates various efficient search algorithms, such as HHBlits and JackHMMER for genetic searching, and HHSearch for template searching.[95]

- The Evoformer Module

The Evoformer module of AlphaFold2, processes inputs (MSA representation and pair representation) from the initial module through deep learning (Figure 2.5). Evoformer generates refined MSA representation and pair representation. The distinctive advantage lies in its ability to interchange information between MSA and pair representations, allowing for mutual

improvement: MSA information can be reinterpreted as pairwise information enhances, and vice versa, facilitating continuous refinement in both aspects.

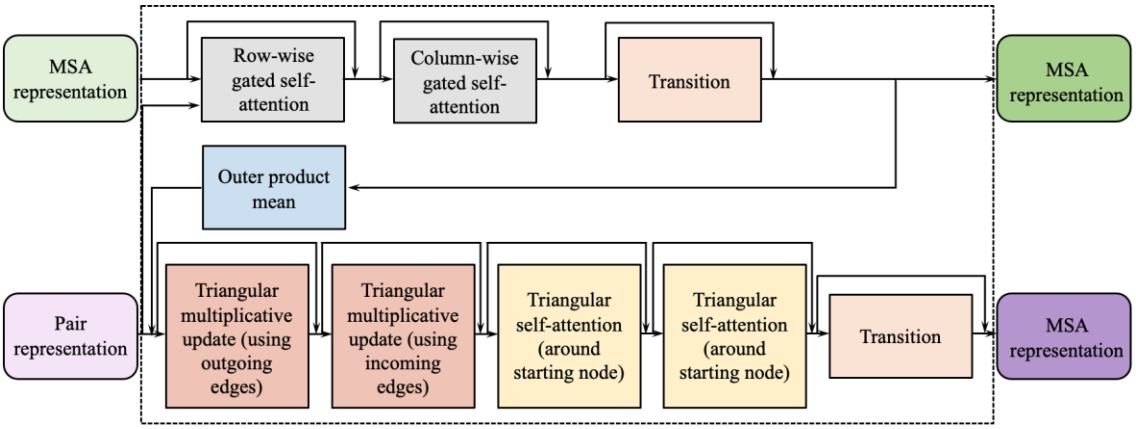


Figure 2.5 The scheme of a block in Evoformer.

The Evoformer architecture comprises 48 blocks with unshared weights, each receiving MSA and pair representations as inputs and producing updated MSA and pair representations as outputs. These representations undergo processing through several layers, incorporating Dropout to mitigate overfitting. Each Evoformer block features two transformer-based pathways with "communication channels" between them. The first pathway operates on the MSA, employing row-wise and column-wise gated self-attention to compute attention over protein symbol matrices efficiently. The second pathway acts on the pair representation, utilizing triangular updates to enforce triangle equivariance, enhancing the network's ability to capture relationships between residues. Both pathways include transition layers for further refinement, employing a 2-layer multilayer perceptron in the MSA pathway to enhance attention mechanisms.

- The Structure Module

The structure module, resembling a decoder, employs a transformer neural network to convert the abstract representation of protein structure into the 3D atom coordinates of target proteins. The structure module operates each residue as an individual and predicts the necessary rotations and translations to accurately position each residue (Figure 2.6).

The structure module of AlphaFold2 comprises three key inputs: a single representation, a pair representation from the Evoformer module, and backbone frames. Each residue is represented as a triangle, with its vertices corresponding to the Cα, N, and C atoms. The module outputs the 3D coordinates for all protein atoms. It consists of eight blocks with shared weights, utilizing Invariant Point Attention (IPA) for updating representations. IPA ensures 3D equivariant attention values, insensitive to global rigid motions. The module predicts relative rotations and translations

of each backbone frame and side-chain χ angles. To address stereochemical constraints, AlphaFold2 applies Amber relaxation without compromising prediction accuracy.

In the last process, AlphaFold2 employs the recycling mechanism three times during iterative refinement for both training and testing (Figure 2.6). This mechanism is widely used in computer

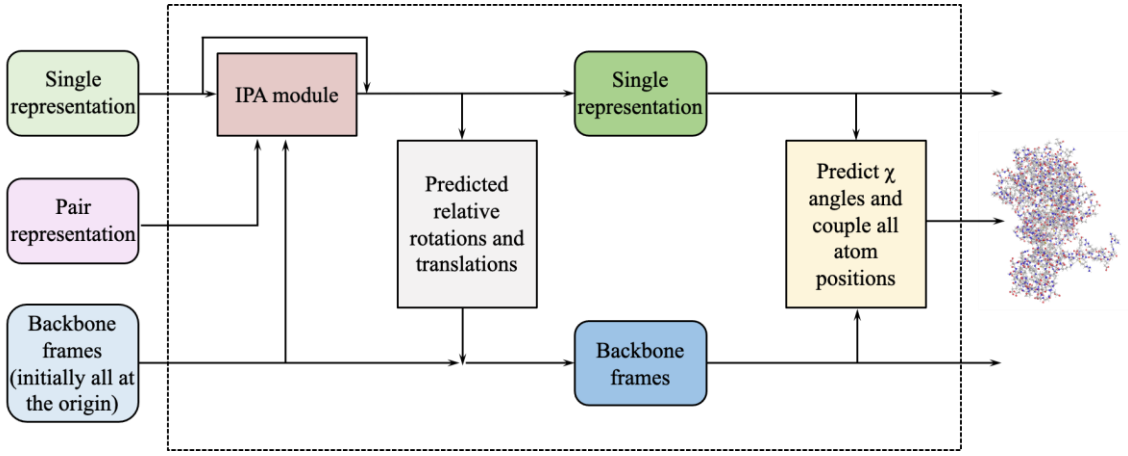


Figure 2.6 Components of a block in the structure module.

vision and enables the network to deepen and process multiple versions of input features without a substantial increase in parameters or training time. In each recycling step, the model integrates previous outputs as additional inputs. AlphaFold2 specifically recycles the predicted backbone atom coordinates from the structure module, as well as the output pair representations and the initial row of MSA representations from the Evoformer.

3 Computational Study on the Catalytic Mechanism and Substrate Specificity of the Aromatic Substrate *C*-Methyltransferase Fur6

3.1 Methods

3.1.1 pK_a Prediction through Epik

PHN and its five analogs which were selected from the biosynthetic pathway of furaquinocin D, namely, 8-AF, FLA, MOP, 8-diazoflaviolin (8-DFA), and 2,4,5,7,8-pentahydroxynaphthalene-1-diazonium (PND), were subjected to pK_a analysis. The pK_a values of all hydroxy groups of PHN and its analogs were calculated with Epik (Release 2021-1)[96,97] using the optimized geometry obtained by Gaussian 16 Rev C.01.[98] Default parameters were used throughout the calculations.

3.1.2 MD Simulations of Fur6 Model Complexed with SAM and PHN

The initial coordinates of Fur6 were obtained from the crystal structure (PDB code: 8HAR; the N-terminal amino acid Met1 in the crystal structure which corresponds to Met22 in NCBI BAE78974). The model structures of the Fur6-SAM-PHN ternary complex were constructed manually. The position of the transferable methyl group of SAM was built by superimposing an optimized SAM structure on that of SAH observed in the crystal structure. The model of PHN was generated in the monoanionic form depending on the pK_a study of PHN. The following presumptions were applied when PHN was docked into the Fur6 structure. First, the overall structure of PHN should be located near the position of the methyl group of SAM, which is necessary for the S_N2-like reaction to initiate. Second, the C3 atom of PHN must be located within 4 Å from the carbon atom of the transferable methyl group of SAM to facilitate the methylation reaction. Third, PHN should be surrounded by several residues that can stabilize the ligand in the active site through hydrogen bonds or π–π stacking.

Since Fur6 forms a dimer structure, MD simulations were performed for dimeric forms. Hydrogen atoms were added to the protein, and all histidine residues were protonated at the Nε2 atom. The geometries of SAM and PHN were optimized at the B3LYP/6-31G(d) level, and their partial charges were obtained by the restrained electrostatic potential (RESP) method using the electrostatic potentials calculated at the HF/6-31G(d) level for the optimized geometries with

Gaussian 16 Rev C.01.[98] The ANTECHAMBER module[99] was used to parameterize the cofactor and the substrate. The ff14SB force field[100] and the general AMBER force field 2 (GAFF2)[101] were used for the protein and organic molecules, respectively. For the MD simulations, the initial coordinates of the methyl group of SAM were generated by the LEaP module[102] using the optimized SAM geometry as a template.

The prepared complex structure was fully solvated with the TIP3P explicit water model[103] in a cubic periodic box with an edge length of 120 Å using the AMBER LEaP module. Na^+ and Cl^- ions were added at a concentration of 0.15 M to neutralize the system. The system was first relaxed using 200 steps of steepest descent minimization with 1,000 kcal mol^{-1} $Å^{-2}$ positional restraints applied to the nonhydrogen atoms of the protein. The restraints were then removed, and the entire system was subjected to 200 steps of steepest descent minimization. Next, MD simulations were performed for 1 ns under the *NPT* ensemble to gradually heat and equilibrate the system at 300 K and 1 atm. In this equilibration run, distance restraints were imposed on atom pairs that form hydrogen bonds in the active site (Table 3.1). After equilibration, 40-ns or 50-ns production runs were performed without distance restraints. The van der Waals interactions were calculated with a cutoff radius of 10 Å. The particle mesh Ewald (PME) method[104] was used to calculate electrostatic interactions. The SHAKE algorithm[105] was used to constrain the bonds, including hydrogen atoms. The integration time step was set to 2 fs. The Berendsen weak coupling algorithm[106] was used to maintain a constant temperature and pressure. All energy minimization, equilibration, and production runs were performed using the PMEMD module of AMBER 20.[107]

Table 3.1 The sets of distance restraints used in MD simulations

Atom 1	Atom 2	Equilibration distance r (Å)	Force constants k (kcal mol^{-1} Å^{-2})
Tyr137-OH	PHN359-O2	3.20	2.0
SAM358-OXT	Thr158-HG1	3.20	2.0
Arg147-NH1	Ala141-O	3.20	2.0
Arg147-NH1	Glu209-O	3.20	2.0
Arg147-NH2	Glu209-O	3.20	2.0
Arg147-NH2	Glu213-OE1	3.20	2.0
Met154-SD	PHN359-C6	4.20	2.0
Trp150-HE1	PHN359-O3	3.20	2.0
Ala181-O	SAM358-N	3.20	2.0

3.1.3 QM/MM Analysis of the Enzymatic Reaction

The initial model of QM/MM analysis was constructed using snapshots extracted from the MD simulation in Gaussian 16 Rev C.01.[98] The activation and reaction energies were calculated using our own *n*-layer integrated molecular orbital and molecular mechanics (ONIOM) method.[94,108-110] A two-layer ONIOM system was built and water molecules more than 5.0 Å away from the protein atoms were removed to reduce the calculation cost. Since the B3LYP/6-31+G(d,p)[111,112] level of theory in the QM region and AMBER ff14SB force field/GAFF2 in the MM region have shown reliable accuracy and computational performance in the previous methyltransferase study,[59] I also chose these calculative levels for this study. The covalent boundary between the QM and MM regions was capped by hydrogen-linking atoms. Considering the effect of hydrogen bonding and calculation cost, Tyr137, Trp150, Asp252, PHN, and a part of SAM were treated as the QM region

(Figure 3.1). The reactant state, intermediate state, and TS were optimized by the ONIOM mechanical embedding scheme.

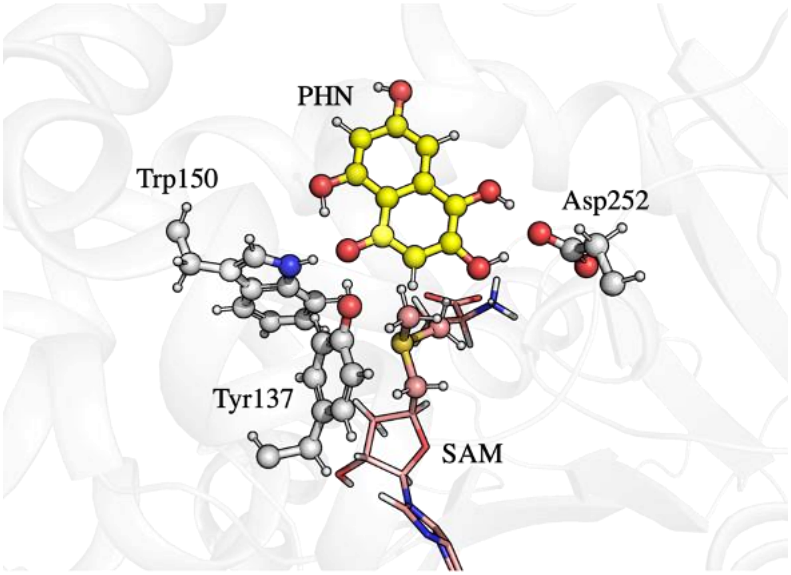


Figure 3.1 QM/MM model used for the ONIOM calculation of the Fur6-SAM-PHN ternary complex. The QM region is shown as a ball-and-stick model, and part of SAM in the MM region is shown as lines. Carbon atoms in PHN, SAM, and Fur6 residues are shown in yellow, pink, and gray, respectively. Oxygen, nitrogen, sulfur, and hydrogen atoms are shown in red, blue, yellow-orange, and white, respectively.

In the initial structural optimization, all the atoms in the models were able to move to reduce steric repulsion. Then, the residues within 7 Å of the QM region were allowed to move; conversely, the remaining part was fixed. The TAO package[113] was applied to build the coordinates, evaluate the ONIOM energies, and update the partial charges of the MM region to reduce the error caused by using fixed charges in the QM region. The structural optimization and the update of partial charges were continued until the decrease in the ONIOM energy was less than 0.1 kcal mol^{-1}. The intrinsic reaction coordinate (IRC) calculations were conducted to confirm that the coordinates of the TS could connect the correct reactant and product structures. The ONIOM electronic embedding (ONIOM-EE) scheme was applied to converged structures to obtain single-point energies. This approach aims to better describe the electrostatic interaction between the QM and MM regions and enable the polarization of the QM wavefunction.

3.1.4 PCM Study on Solution-phase Free Energy (G_{soln})

The geometries of the intermediate and product were extracted from QM/MM models, and then the optimization and frequency calculations were carried out at the B3LYP/6-311G(d) level of theory. The optimized structures were then used to calculate the solution- and gas-phase-free

energies, respectively, and the solution was set as water. Patel and Masunov reported that M05-2X/6-31G(d) showed higher accuracy than other theory levels in the PCM model, so this level of theory was chosen this time.[114] The procedure and equations for the calculation of G_{soln} were referred to in Ho et al.[115] All calculations were carried out in Gaussian 16 Rev C.01.[98]

3.1.5 Calculation of the Potential Energy Barrier *in vacuo*

Energy barriers of the methylation reactions on the PHN analogs were estimated using a truncated SAM model. The PHN structure and the -$CH_2S^+(CH_3)$ CH_2- part of the SAM structure were extracted from the QM/MM model, and the terminals of the truncated SAM model were capped with hydrogens (denoted as SAM-1, Figure 3.2). The coordinates of SAM-1 were reserved, while the coordinates of PHN were replaced by the analogs' coordinates. The structures were optimized at the B3LYP/6-31G(d) level of theory. The reaction coordinate was defined as the distance between two reacting atoms: the carbon atom in the methyl group of SAM-1 and the corresponding methylation center of PHN and its analogs, for example, the C3 atom of PHN in the PHN reaction. For the reaction between SAM-1 and MOP, considering the symmetry of MOP and the similarity of MOP to PHN, I also set the reaction coordinate between the C6 atom of MOP

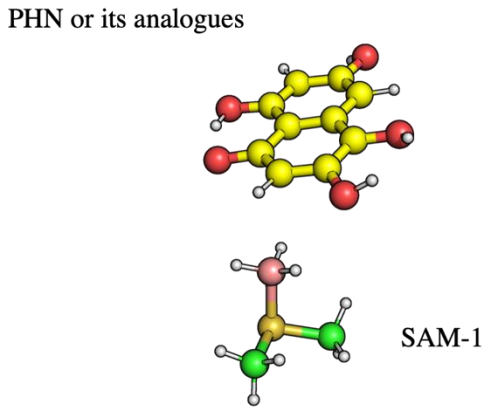


Figure 3.2 Model used for the potential energy barrier calculation. The position of PHN represents itself and its analogs. The carbon atom in the transferrable methyl group is shown in pink, and other carbon atoms in PHN and SAM-1 are shown in yellow and green, respectively. Oxygen, sulfur, and hydrogen atoms are shown in red, yellow-orange, and white, respectively.

and the transferrable methyl group of SAM-1 to explore methylation. The same level of theory was used in potential energy surface (PES) scans, the number of steps was 15, and the step size was 0.1 Å. These calculations were performed using Gaussian 16 Rev C.01.[98] The difference in the energies between the initial state and the step with the largest energy was used as the estimated value of the energy barrier.

3.1.6 Calculation of the HOMO Energy and Its Analogs

The geometries of PHN and its analogs were determined using the structures built for calculating the potential energy barriers *in vacuo*. The energy calculations were carried out at the HF/6-31G(d) level of theory through Gaussian 16 Rev C.01.[98]

3.1.7 pK_a Prediction through PROPKA

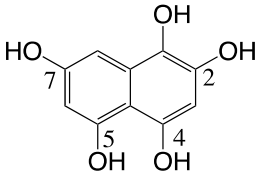
The pK_a values Tyr137 in apoprotein and holoprotein were calculated with the PROPKA[116] package in Protein Preparation Wizard of Schrödinger, and the default parameters were used throughout the calculations.

3.2 Results and Discussion

3.2.1 Elucidation of the Fur6 Model Complexed with SAM and PHN

The pK_a values of all hydroxy groups of PHN (Table 3.2) calculated by using the software program Epik (Release 2021-1)[96,97] suggested that PHN should react in monoanionic form, and the hydroxy group at C5 was deprotonated. Therefore, PHN was prepared in C4-OH, C5-O^-.

Table 3.2 pK_a values of all hydroxy groups of PHN

Structure	Position	pKa
	1-OH	9.13
	2-OH	8.02
	4-OH	6.80
	5-OH	6.74
	7-OH	9.27

MD simulation revealed a stable binding mode of PHN in Fur6 (Figure 3.3). The C3 atom of PHN in the complex was located at an average distance of 3.41 ± 0.19 Å from the carbon atom of the methyl group of SAM (Figure 3.4), which is sufficiently close for the methylation reaction. Tyr137, Trp150, and Asp252 form hydrogen bonds with PHN, and the aromatic moieties of Tyr98, His249, Phe297, Phe298, Phe301, and Phe302 appear to stabilize the ligand in the active site.

When these residues were mutated, the methylation activity of Fur6 disappeared,[34] suggesting that PHN is recognized by Tyr137, Trp150, and Asp252 in the Fur6 active site. Thus, I conducted

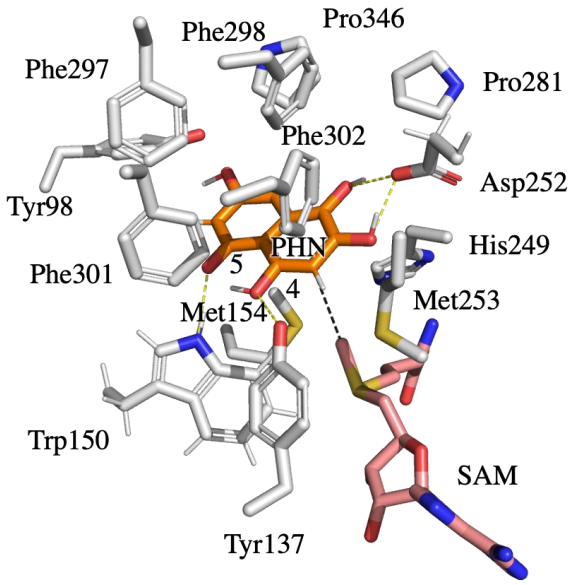


Figure 3.3 MD simulations of the Fur6 model complexed with SAM and PHN. Active site structures of Fur6-SAM-PHN (C4-OH, C5-O^-).

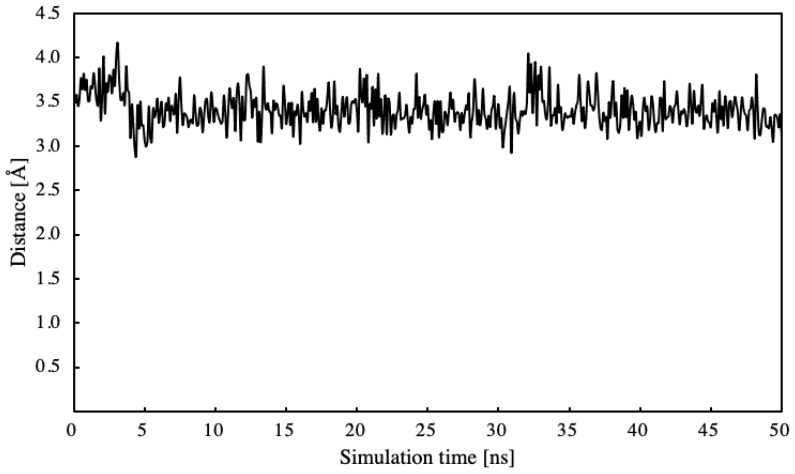


Figure 3.4 MD simulations of the Fur6 model complexed with SAM and PHN. Plot of the distance between the methyl group of SAM and C3 of PHN, taken from the unstrained 50-ns MD simulation of the Fur6-SAM-PHN (C4-OH, C5-O^-) complex.

a QM/MM analysis by using the structure from the MD simulation. However, after the initial structural optimization, the proton of the C4-hydroxy group was transferred to the C5-oxygen anion, which changed the reactant from C4-OH, C5-O^- to C4-O^-, C5-OH. Considering this change, I performed the MD simulation of Fur6 complexed with SAM and PHN (C4-O^-, C5-OH).

By comparing MD simulations of Fur6-SAM-PHN (C4-O$^-$, C5-OH) (Figure 3.5) with Fur6-SAM-PHN (C4-OH, C5-O$^-$), the average distance between the C3 atom of PHN and the transferrable methyl group of SAM was 3.34 ± 0.17 Å (Figure 3.6), which is closer than the MD

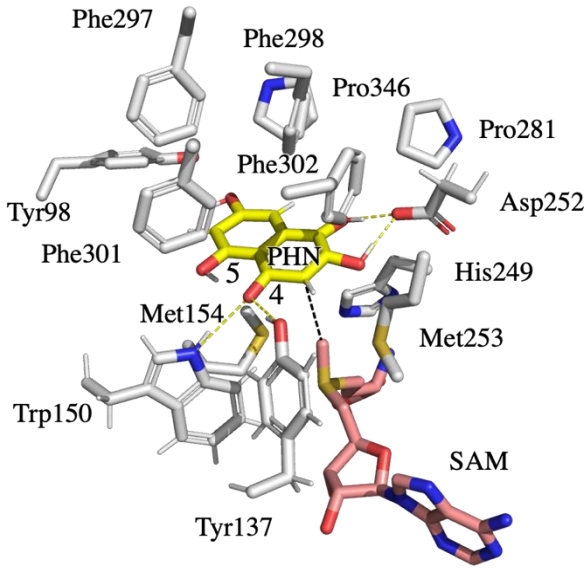


Figure 3.5 MD simulations of the Fur6 model complexed with SAM and PHN. Active site structures of Fur6-SAM-PHN (C4-O$^-$, C5-OH).

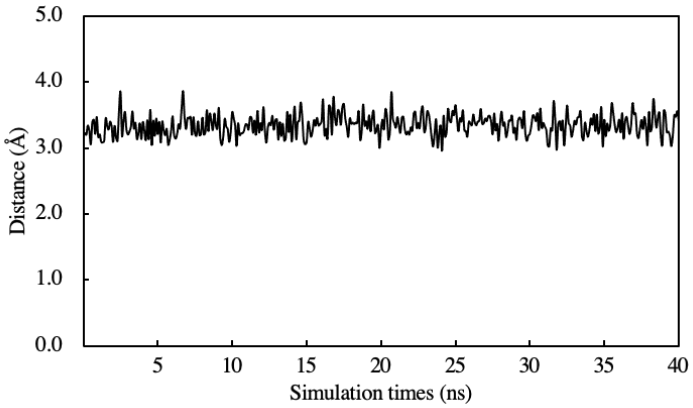


Figure 3.6 MD simulations of the Fur6 model complexed with SAM and PHN. Plot of the distance between the methyl group of SAM and C3 of PHN, taken from the unstrained 40-ns MD simulation of the Fur6-SAM-PHN (C4-O$^-$, C5-OH) complex.

simulation result of Fur6-SAM-PHN (C4-OH, C5-O$^-$) complex. The decrease in the distance would help the reacting atoms interact much more easily. Although the calculated pK_a value of the hydroxy group on the C4 atom (6.80) was slightly larger than that on C5 (6.74), the MD

simulations of Fur6-SAM-PHN (C4-O$^-$, C5-OH) complex showed that the rearrangement of the proton to the C5 atom is reasonable and will efficiently facilitate the methylation reaction. Therefore, I decided to use the Fur6-SAM-PHN (C4-O$^-$, C5-OH) model in further QM/MM studies.

3.2.2 QM/MM Study on the Methyl Transfer Step Catalyzed by the Fur6

To further investigate the methyl transfer step, I found the transition state (TS) of the reaction and calculated the activation ($\Delta E^{\ddagger}$) and reaction energies (ΔE) using the ONIOM method.

The ONIOM calculations showed that the methyl group of SAM was transferred to PHN through a transition state (TS1, Figure 3.7 A) to reach the intermediate state (Figure 3.8 A-C). In the reactant state (Figure 3.8 A), the distance between the C3 atom of PHN and the methyl group of SAM was 2.97 Å. In TS1 (Figure 3.8 B), the methyl group became almost planar and moved closer to the C3 atom of PHN at 2.53 Å. Moreover, the C3 atom of PHN, methyl group, and sulfur atom of SAM stayed in an almost linear configuration (177.48°), which is consistent with the S_N2 reaction mechanism in other methyltransferases.[59,117] In the intermediate state (Figure 3.8 C), the methyl group covalently bonded to the C3 atom of PHN, while SAM converted to SAH. The calculated $\Delta E^{\ddagger}$ of TS1 from the reactant state was 8.6 kcal mol^{-1} (Figure 3.8 D), and the calculated ΔE of the intermediate state was –32.7 kcal mol^{-1}. Although I obtained the intermediate state, the extra proton was still bound to the C3 atom of PHN. I further investigate the subsequent deprotonation step in the next section.

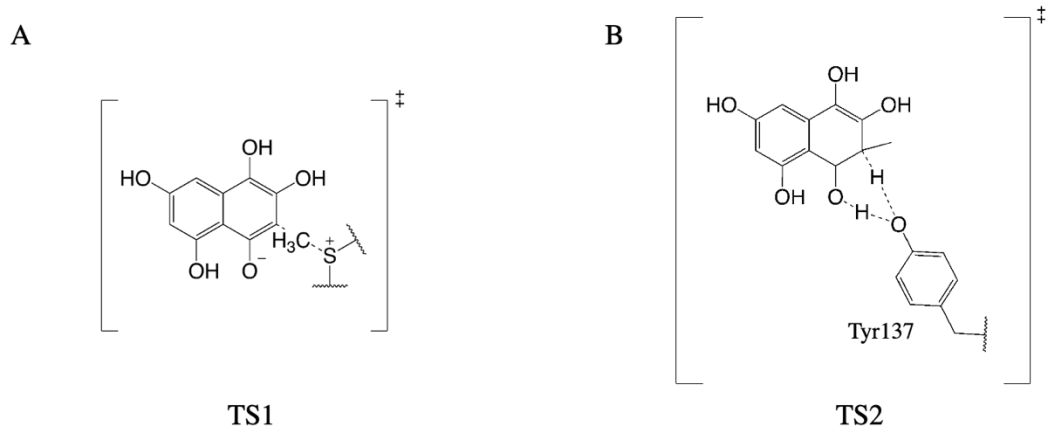


Figure 3.7 TS structure calculated by QM/MM analysis. A, B, the transition state structure of the methyl transfer step (A), the transition state structure of the deprotonation step (B).

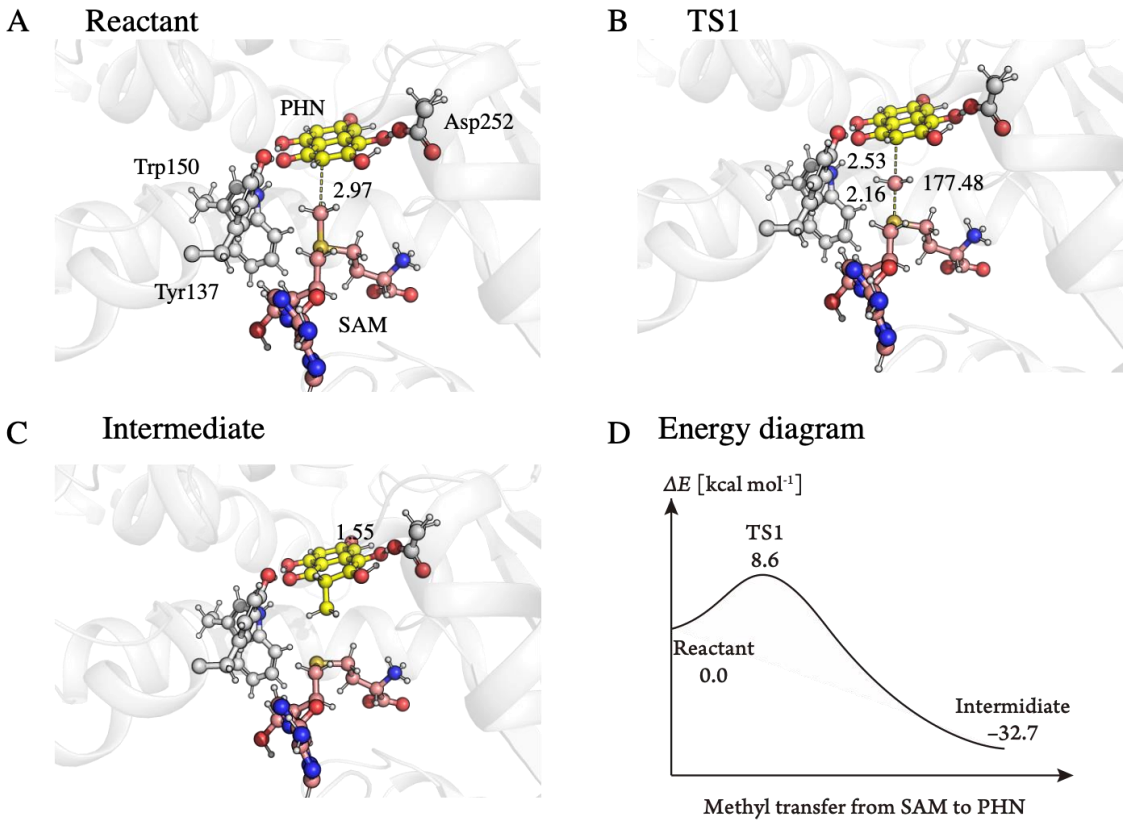


Figure 3.8 QM/MM analysis of the methylation reaction catalyzed by Fur6. A–C) Structures of the reactant (A), TS1 (B), and intermediate (C). Bond lengths are shown in angstroms, and the angle is shown in degrees. (D) Energy diagram of methyl transfer from SAM to PHN.

3.2.3 QM/MM and PCM Study on the Deprotonation Step

To clarify the deprotonation pathway of an extra proton left on the C3 atom of PHN, I first performed a QM/MM analysis with the intermediate state. I hypothesized that Tyr137 would assist the deprotonation step by passing the extra proton, similar to Tyr134 in the S*s*-SibL *C*-methylation mechanism reported by Chen et al.[118]

I obtained the intermediate TS2 (Figure 3.7 B) and product structures from IRC calculations and structural optimizations (Figure 3.9). In the TS2 state, the extra proton moved toward the hydroxy group of Tyr137, while the proton of the hydroxy group of Tyr137 moved toward the oxygen anion of the C4 atom of PHN simultaneously. The final product is neutral 3-Me-PHN. However, the calculated $\Delta E^{\ddagger}$ of TS2 from the intermediate state was 43.9 kcal mol^{-1}, and the calculated ΔE of the product state was –34.3 kcal mol^{-1}, indicating that the deprotonation reaction should not occur with the help of Tyr137. Moreover, I also discussed the reactive form of Tyr137

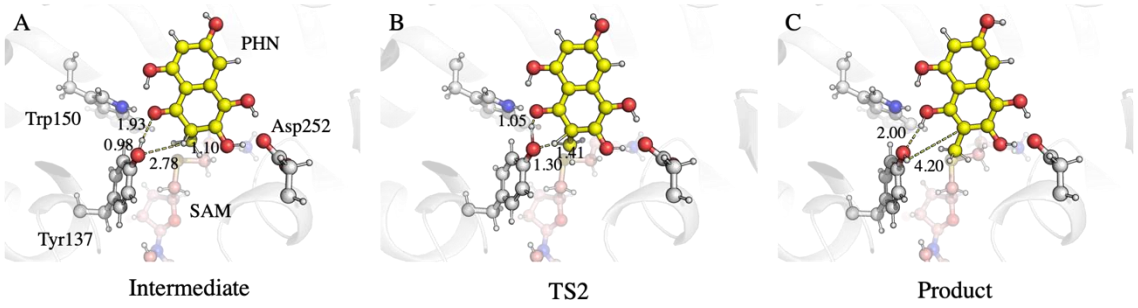


Figure 3.9 QM/MM analysis of the deprotonation mechanism catalyzed by Fur6. A–C, the optimized structure of the Intermediate (A), TS2 (B), and Product (C) states for deprotonation in Fur6. The bond lengths are shown in angstroms. Carbon atoms in PHN, SAM, and Fur6 residues were shown in yellow, pink, and gray, respectively. Oxygen, nitrogen, sulfur, and hydrogen atoms were shown in red, blue, yellow-orange, and white, respectively.

to elucidate whether the reactive form of Tyr137 influences the $\Delta E^{\ddagger}$ of TS2 from the intermediate state. The predicted pK_a value of the hydroxy group of Tyr137 in apoprotein and holoprotein is 12.99 and 13.18, respectively. The pK_a value indicated that Tyr137 should react in the neutral form rather than a monoanionic form. Therefore, I had to consider another possible pathway for proton transfer. Because I did not find any residues near the C3 atom of PHN that could act as a base to remove the extra proton (Figure 3.5), I hypothesize that a water molecule comes and assists in completing the deprotonation reaction. The deprotonation step was evaluated by calculating the pK_a value of the extra proton. To calculate more accurate pK_a values of the intermediates considering the effect of PCM solvation, I employed a calculation method developed by Liptak and Shields.[119] The calculation results showed that the estimated pK_a value using the method of Lipton and Shields is 5.20, which means that the extra proton of the intermediate will easily dissociate and the dissociated equilibrium between the intermediate and product will move toward the product in solvent. Furthermore, I investigated the active site of intermediate state of Fur6 complex model through performing the 40-ns MD simulation for three times. The probability of water molecules appeared near the C3 atom of PHN is 0.8 %, the low probability suggesting that deprotonation could not occur in the active site. Therefore, based on the results of a combination of QM/MM analysis, and PCM solvation methods, I concluded that deprotonation presumably occurs in solvent after the intermediate is ejected from Fur6.

3.2.4 Specificity of Methylation Catalyzed by Fur6

In the study on the biosynthetic pathway of furaquinocin D, the activity of Fur6 was evaluated in the sequential reaction of Fur5 and Fur6 using 8-AF as the substrate. 8-AF is converted into 8-DFA by Fur5, and 8-DFA is further converted into PHN via PND. PHN is easily oxidized into

FLA. Interestingly, only the methylation product of PHN was detected in the sequential reaction despite the presence of these structurally analogous compounds in the reaction mixture.[34]

To investigate the origin of the specificity of methylation catalyzed by Fur6, I selected five PHN analogs (8-AF, FLA, MOP, 8-DFA, and PND) and predicted their reactivities by predicting protonation states, energy barriers of the methylation reaction, and HOMO energies. MOP was added to examine the effect of the substituent of C6. The protonation state of each analog was predicted by calculating the pK_a values of all hydroxy groups in the analogs (Table 3.3-Table 3.7).

Table 3.3 pK_a values of all hydroxy groups of 8-AF

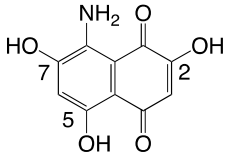
Structure	Position	pK_a
	2-OH	5.73
	5-OH	10.27
	7-OH	5.37

Table 3.4 pK_a values of all hydroxy groups of FLA

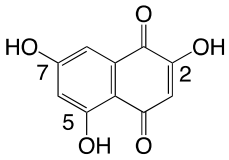
Structure	Position	pK_a
	2-OH	4.08
	5-OH	9.06
	7-OH	6.88

Table 3.5 pK_a values of all hydroxy groups of MOP

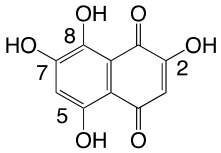
Structure	Position	pK_a
	2-OH	3.90
	5-OH	9.58
	7-OH	5.95
	8-OH	6.03

Table 3.6 pK_a values of all hydroxy groups of 8-DFA

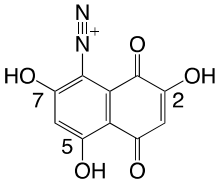
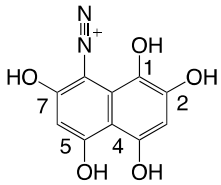
Structure	Position	pK_a
	2-OH	2.38
	5-OH	6.01
	7-OH	−1.16

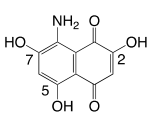
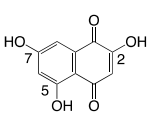
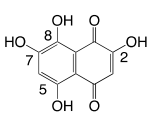
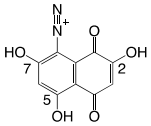
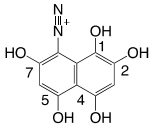
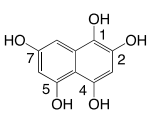
Table 3.7 pK_a values of all hydroxy groups of PND

Structure	Position	pK_a
	1-OH	9.11
	2-OH	6.07
	4-OH	3.34
	5-OH	3.04
	7-OH	0.62

Since all the analogs have pK_a values less than 7.0, I proposed that they would react in their monoanionic forms rather than in their neutral forms. Hence, I estimated the energy barriers for the monoanionic forms of PHN and its analogs (Table 3.8).

The analogs that can be catalyzed by Fur6 should satisfy the following two requirements. First, they should form appropriate interactions with the enzyme in the active site. Since PHN forms intermolecular hydrogen bonds with Tyr137, Trp150, and Asp252, these interactions should be preserved to stabilize the substrate in the active site (Figure 3.5). It should be noted that a protonated hydroxy group is necessary to form a hydrogen bond with Asp252, whereas both protonated and deprotonated forms are allowed to interact with Tyr137 and Trp150. Second, the energy barrier of the reaction should be sufficiently low to allow the reaction to occur at room temperature (< 20 kcal mol^{-1}). To reduce the computational cost, I estimated the energy barriers of the analogs *in vacuo* (*i.e.*, without atoms in protein) using a truncated model of SAM. Given the small difference for PHN between the energy barrier calculated with the QM/MM method (8.6 kcal mol^{-1}) and that calculated *in vacuo* (10.4 kcal mol^{-1}), I assume that large differences would not occur for the other substrates. Moreover, I deduce the reason that the potential energy barrier calculated *in vacuo* is slightly higher than in the enzyme because it ignores the effect of residues in the active site.

Table 3.8 Summary of calculation on pK_a value, potential energy barrier, and HOMO energy

	8-AF	FLA	MOP	8-DFA	PND	PHN
Structure						
pK_a (position; smallest value)	7-OH; 5.37	2-OH; 4.08	2-OH; 3.09	7-OH; −1.16	7-OH; 0.62	4-OH; 6.80
Potential energy barrier (kcal mol^{-1})	21.1	13.7	13.5 (C3) 52.3 (C6)	31.1	19.8	10.4
HOMO energy (eV)	−2.83	−3.49	−3.43	−8.71	−7.77	−1.86

The results of the pK_a calculations suggested that FLA and MOP would lose the proton of the C2-hydroxy group and that the hydrogen bond formed with Asp252 should be broken, making it difficult for these molecules to stably bind in the active site. Because MOP can form hydrogen bonds with Asp252 via the C7- and C8-hydroxy groups, I also estimated the energy barrier of the C6-methylation reaction. As the C2-hydroxy group of 8-AF and 8-DFA and the C1- and C2-hydroxy groups of PND were predicted to be protonated, these three analogs should bind in the active site, as does PHN. However, the estimated energy barriers of 8-AF, 8-DFA, and PND were higher than that of PHN by more than 9 kcal mol^{-1}, suggesting that Fur6 would only catalyze methylation against these analogs at a much slower reaction rate compared to that of PHN. The energy barrier of the C6-methylation reaction of MOP was 52.3 kcal mol^{-1}, indicating that the reaction could not occur at room temperature. Moreover, I compared the energy levels and the molecular orbital coefficients of HOMO between PHN and its five analogs. I found that PHN possessed the highest HOMO energy (−1.86 eV) (Table 3.8) with sufficiently large coefficients for the C3 atomic orbitals (Table 3.9), which suggests that PHN is the most reactive with the methyl cation. Thus, my results can explain the experimental observation, in which no other

methylation products were detected except 3-Me-PHN, and strongly support the conclusion of the previous study that PHN is the physiological substrate of Fur6.

Table 3.9 Coefficients of HOMO calculation of PHN and its analogs

Compound	Coefficient of 2PZ orbital of C3	Coefficient of 3PZ orbital of C3
8-AF	0.01	0.01
FLA	0.39	0.42
MOP	0.39	0.41
8-DFA	0.08	0.07
PND	−0.20	−0.20
PHN	−0.25	−0.30

3.3 Conclusion

In this study, I systematically investigated the reaction mechanism and substrate specificity of Fur6. First, through the MD simulation and QM/MM analysis, I found that the natural substrate PHN should react in the C4-O$^-$, C5-OH form rather than the C4-OH, C5-O$^-$ form. Moreover, Tyr137, Trp150, and Asp252 are important for the reactivity of Fur6, as they form hydrogen bonds with the substrate. Second, by using the QM/MM analysis and free energy calculation with the PCM method, I clarified the entire reaction mechanism of Fur6 (Figure 3.10). The methylation follows the S_N2 mechanism, similar to other known methyltransferases.[59,117] The nucleophilic

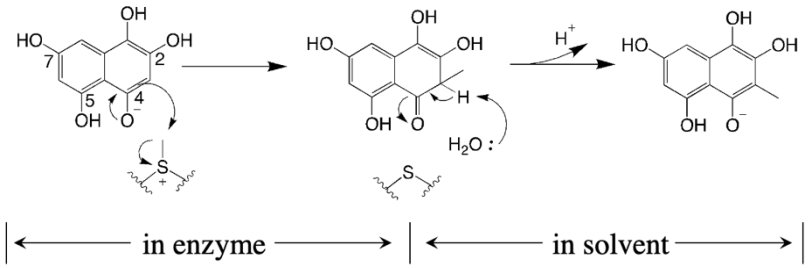


Figure 3.10 Proposed mechanism for the C3 methylation reaction of PHN catalyzed by Fur6.

attack from the C3 atom on the methyl group of SAM produces SAH and an intermediate in which the C3 atom is methylated, with a $\Delta E^{\ddagger}$ of approximately 8.6 kcal mol^{-1}. The methylated intermediate is then rearomatized after it is presumably ejected into the water solvent. Finally, pK_a prediction and QM calculation suggest that among the analogs selected from the biosynthetic

intermediates of furaquinocin D, although 8-AF, 8-FDA, and PND form similar intermolecular interactions to that of PHN, their higher potential energy barriers result in a lower reaction rate with Fur6 compared to PHN, or they cannot even react. Moreover, these results also computationally demonstrate that PHN should be the authentic substrate of Fur6.

Methyltransferase is a well-known biosynthetic enzyme in the study of natural products, as it can enhance the bioactivity of a large number of natural products.[120,121] A deeper understanding of the reaction mechanism and continuous exploration of the nature of substrate specificity will assist in producing "unnatural" natural products, which possess the desired functionality based on therapeutic and other needs. This study broadens my horizon in the field of methyltransferases and will contribute to further enzyme engineering studies.

4 Computational Study on the Catalytic Mechanism of the Aromatic Prenyltransferase Fur7

4.1 Methods

4.1.1 pK_a Prediction through Epik

pK_a prediction for all hydroxy groups of MMT and MMF was carried out using Epik (Release 2021-1)[96,97] with the optimized geometry obtained by Gaussian 16 Rev C.01.[98] The procedures were the same as Section 3.1.1.

4.1.2 Calculation of the HOMO energy of MMT and MMF

The geometries of MMT and MMF were built in GaussView 6.1.1. The structural optimizations and energy calculations were carried out at the same level of theory as Section 3.1.6.

4.1.3 MD Simulations of Fur7 Model Complexed with GPP and MMT

The initial coordinates of Fur7 were obtained from the predicted structure generated by AlphaFold2 (version 2.1). The amino acid sequence was provided by Dr. Noguchi, T., and matches exactly with the sequence published in NCBI (accession code: Q2L6E3.1).

- The Procedure of the MD simulation of Fur7-GPP-MMT(5-O$^-$)

The model structures of the Fur7-GPP-MMT ternary complex were constructed manually. The position of the GPP was built by superimposing an optimized GPP structure on that of GSPP observed in the NphB crystal structure (PDB code: 1ZB6). The model of MMT was generated in the monoanionic form depending on the pK_a study. The following presumptions were applied when MMT was docked into the Fur7 structure. First, the overall structure of MMT should be located near the position of the geranyl group of GPP, which is necessary for the reaction to initiate. Second, the C6 atom of MMT must be located within 4.5 Å from the C3 atom of GPP to facilitate the prenylation reaction. Third, MMT should be surrounded by several residues that can stabilize the ligand in the active site through hydrogen bonds, especially π–π stacking.

The approach of creating the parameter file of GPP and MMT is identical to that for SAM and PHN, as described in Section 3.1.2. The setting of MD simulations is also the same as Section

3.1.2, except for the distance restraints that were imposed on atom pairs that form hydrogen bonds in the active site (Table 4.1).

Table 4.1 The sets of distance restraints used in MD simulations

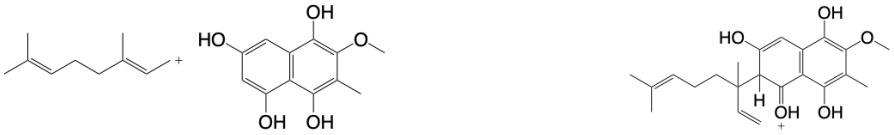
Atom 1	Atom 2	Equilibration distance *r* (Å)	Force constants *k* (kcal mol^{-1} $Å^{-2}$)
Phe215-CE2	MMT309-C1	4.20	2.0
Glu229-OE2	Arg59-NH1	3.20	2.0
Glu229-OE2	Arg230-NH1	3.20	2.0
Lys121-NZ	Asp169-OD2	3.20	2.0
Glu277-OE2	Asn57-ND2	3.20	2.0
Asp112-OD2	Arg53-NH1	3.20	2.0
Lys236-NZ	Asn250-O	3.20	2.0
Gln265-OE1	Lys236-NZ	3.20	2.0
Glu238-OE1	Gln265-NE2	3.20	2.0
Lys120-NZ	Asp63-OD2	3.20	2.0
Lys236-NZ	Trp288-CZ2	4.20	2.0

- The Procedure of the MD Simulation of Fur7-GPP-MMT (neutral form)

The model structures of the Fur7-GPP-MMT ternary complex were constructed manually by replacing the MMT (5-O^-) with MMT (neutral form). The process of creating the parameter file of MMT (neutral form) and setting up the MD simulations is the same as described in Section 3.1.2, except that the production runs were set for 30 ns.

- The Procedure of the MD Simulation of Fur7-OPP^{4-}-intermediate 2 (IM2)

For convenience, the state after the C-O bond cleavage was denoted as intermediate 1 (IM1), and the state of the bond formation between C3-GPP and C6-MMT as denoted as IM2 (Figure 4.1). The model structures of the Fur7-OPP^{4-}-IM2 ternary complex were manually constructed.



IM1 IM2

Figure 4.1 Structures of IM1 and IM2.

The position of the OPP^{4-} was created by referencing the position of OPP^{4-} in GSPP observed in the NphB crystal structure, and slightly adjusted to improve interaction with the surrounding residues. IM2 was docked to Fur7 in a position surrounded by several residues that can stabilize the ligand in the active site through hydrogen bonds, especially π–π stacking.

The process of creating the parameter file of OPP^{4-} and IM2 was identical to that for SAM and PHN, as described in Section 3.1.2. The parameters and procedures of MD simulations were the same as Section 3.1.2, except the production runs were performed in 10 ns.

- The Procedure of the MD Simulation of Fur7-OPP^{4-}-IM1

The model structure of the Fur7-OPP^{4-}-IM1 ternary complex utilized the structure obtained from the optimized IRC calculation of the QM/MM calculation. The process of creating the parameter file and setting up the MD simulations was the same as described in Section 3.1.2, except that the production runs were set for 10 ns.

4.1.4 QM/MM Analysis of the Enzymatic Reaction

The procedure for preparing input files and performing calculations for QM/MM analysis remains the same as described in Section 3.1.3. This section will only focus on the distinct settings used. Given that the M06-2X/6-31G(d,p) level of theory for the QM region and the AMBER ff14SB force field/GAFF2 in the MM region have demonstrated reliable accuracy and

computational performance in previous studies on prenyltransferase,[66,122] I adopted these calculative levels for this study as well.

- QM/MM Model of Fur7-GPP-MMT(5-O^-)

Considering the effect of hydrogen bonding and calculation cost, Arg53, Arg67, Tyr178, Lys281, Arg230, MMT, GPP, and a water molecule were treated as the QM region (Figure 4.2).

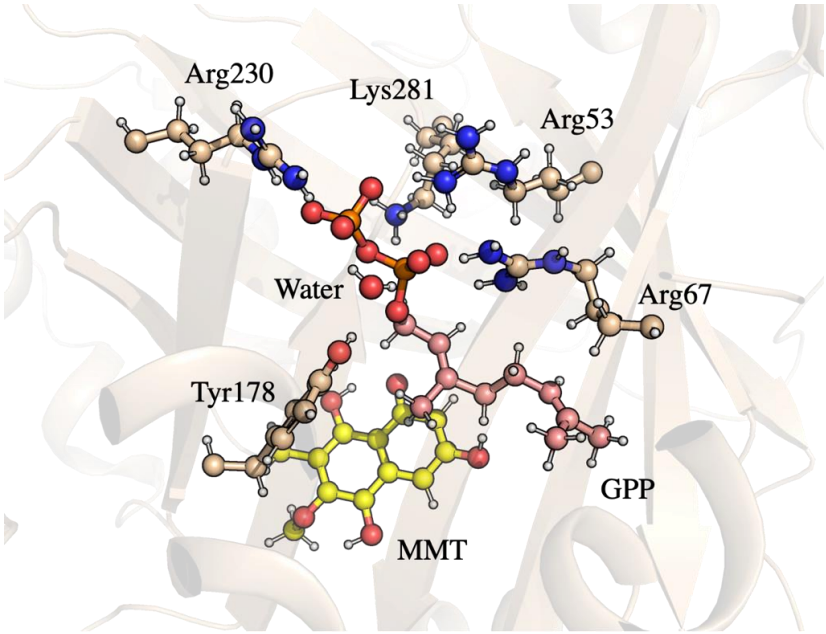


Figure 4.2 QM/MM model used for the ONIOM calculation of the Fur7-GPP-MMT ternary complex. The QM region is shown as a ball-and-stick model. Carbon atoms in MMT, GPP, and Fur7 residues are shown in yellow, pink, and tints, respectively. Oxygen, nitrogen, and hydrogen atoms are shown in red, blue, and white, respectively.

- QM/MM Model of Fur7-OPP^{4-}- IM2

Considering the impact of hydrogen bonding and computational cost, Trp123, Tyr125, and Tyr269 were designated as the QM region (Figure 4.3).

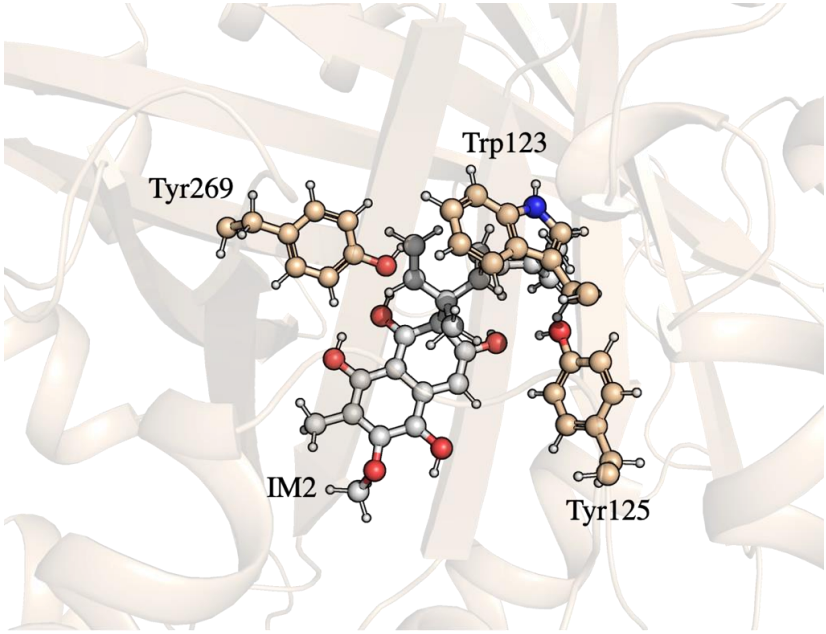


Figure 4.3 QM/MM model used for the ONIOM calculation of the Fur7-OPP^{4-}- IM2 ternary complex. The QM region is shown as a ball-and-stick model. Carbon atoms in IM2 and Fur7 residues are shown in white and tints, respectively. Oxygen, nitrogen, and hydrogen atoms are shown in red, blue, and white, respectively.

4.1.5 2D PES Scan

The 2D PES was generated using the Fur7-GPP-MMT model, and the M06-2X/6-31G(d,p) level of theory was applied to describe the QM layer. The bond length of O1–C1 of GPP and the distance between C3-GPP and C6-MMT were selected as the two reaction coordinates. To create grid points for energy calculations in the ONIOM scheme, the first coordinate was scanned with a step size of 0.44 Å, ranging from 1.53 Å to 3.29 Å, while the second coordinate ranged from 3.42 Å to 1.42 Å with a step size of −0.50 Å. A coarse reaction surface was constructed by connecting a total of 25 energy values calculated from optimizations, during which the entire geometry was relaxed, except for the length of O1–C1 of GPP and the distance between C3-GPP and C6-MMT, as specified by the grid points. All calculations were carried out in Gaussian 16 Rev C.01.[98]

4.1.6 Analysis of the Reaction Pathway of C6-Normal and Reverse Intermediate

Energy barriers of the prenylation reactions, which produce the normal product or reverse product on the C6 of MMT, were estimated using an intermediate model *in vacuo*. In this model, the extra hydrogen at C6 has not deprotonated (Figure 4.4). Subsequently, the retro-biosynthetic analysis was carried out to analyze the reaction pathways of C6-normal and C6-reverse products. The intermediate structures were constructed using GaussView 6.1.1 and optimized at the B3LYP/6-31G(d) level of theory. The reaction coordinate was defined as the distance between two reacting atoms: the carbon atom in the geranyl group and the C6 atom of MMT. The same level of theory was applied in PES scans, with 11 steps and step sizes of 0.15 Å and 0.14 Å for normal and reverse product reactions, respectively. Additionally, IRC calculations were performed to confirm that the coordinates of the transition state connected the correct reactant and product structures. These calculations were conducted using Gaussian 16 Rev C.01.[98]

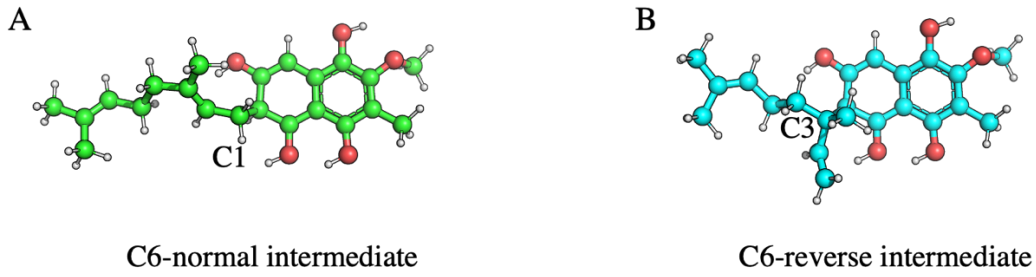


Figure 4.4 Model used for the potential energy barrier calculation. C6-normal product (A), C6-reverse product (B). The carbon atom in the C6-normal product and C6-reverse product are shown in green and cyan, respectively. Oxygen and hydrogen atoms are shown in red and white, respectively.

4.2 Results and Discussion

4.2.1 Analyzation of the Original Substrate of Fur7

As briefly introduced in Section 1.1.3, Kumano, T. et al. originally proposed MMF as the substrate of Fur7.[20] However, Noguchi, T. et al. later discovered that the prenylation was significantly enhanced with the addition of the reductant dithionite. This novel observation led to a hypothesis that the authentic substrate of Fur7 is MMT. However, the product, 6-prenyl-2-methoxy-3-methyl PHN, is easily oxidized, and only the oxidized C6-product was detected.[34] Given the complexity of this observation, I employed a HF-based calculation to estimate the HOMO energy of MMT and MMF.

By comparing the energy levels and the molecular orbital coefficients of HOMO between MMT and MMF, I determined that MMT possesses a higher HOMO energy (−0.26 eV) with sufficiently larger coefficients for the C6 atomic orbitals (Table 4.2). These calculation results indicate that MMT is more reactive for the prenylation reaction occurring on the C6 atom. Consequently, my findings provide an explanation for the experimental observation, where MMF, in the presence of the reductant, exhibits enhanced reactivity compared to its absence. This strongly supports the conclusion reported by Noguchi, T. et al. that MMT is the physiological substrate of Fur7.

Table 4.2 HOMO energy and coefficients of HOMO calculation of MMT and MMF

Compound	HOMO energy (eV)	Coefficient of 2PZ orbital of C6	Coefficient of 3PZ orbital of C6
MMT	−0.26	−0.14	−0.14
MMF	−0.33	0.05	0.04

4.2.2 The 3D model of Fur7 Predicted by AlphaFold2

The 3D structure of Fur7, as predicted by AlphaFold2, exhibits a high level of confidence (Figure 4.5), which was estimated using a per-residue confidence metric called the predicted local

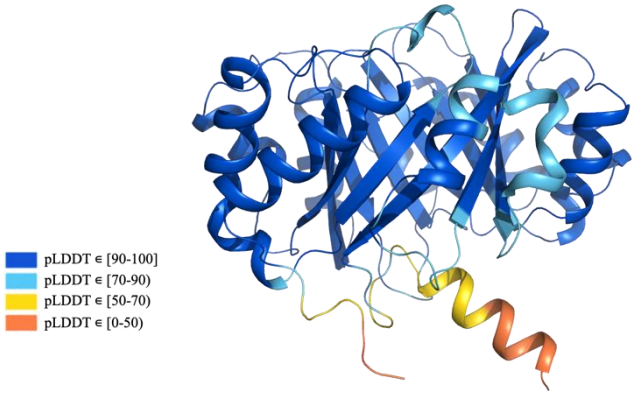


Figure 4.5 The structure predicted by AlphaFold2; the prediction is colored by model confidence band.

distance difference test (pLDDT), generated by AlphaFold.[88] The pLDDT is corresponds to the model's predicated score on the lDDT-Cα metric. The pLDDT scale ranges from 0 to 100, with a high accuracy cutoff typically set at pLDDT > 90, depicted in blue, while a lower cutoff of pLDDT > 70 corresponds to a generally correct backbone prediction, indicated in aquamarine (Figure 4.5). In essence, a lower pLDDT value indicates lower prediction accuracy. In the case of

Fur7, the functional domains exhibit high confidence, whereas only the N-terminal and C-terminal unstructured regions exhibit lower confidence.

I further compared the predicted 3D structure of Fur7 with its homologs, such as NphB and CloQ (Figure 4.6). Fur7 exhibits a similar 3D structure to NphB and CloQ, all of which share a 10-stranded antiparallel β-barrel fold comprised of five repetitive αββα elements, commonly referred to as the "prenyltransferase barrel fold." This similarity suggests that Fur7 belongs to the ABBA-type superfamily, as introduced in Section 1.1.3. Furthermore, several hydrophobic residues located in the "bottom" of the prenyltransferase barrel of Fur7, which is the same as NphB and CloQ, sequester the geranyl tail of GPP or DMAPP. In contrast, the polar residues located at the "top" of the barrel may aid in the binding of the diphosphate.

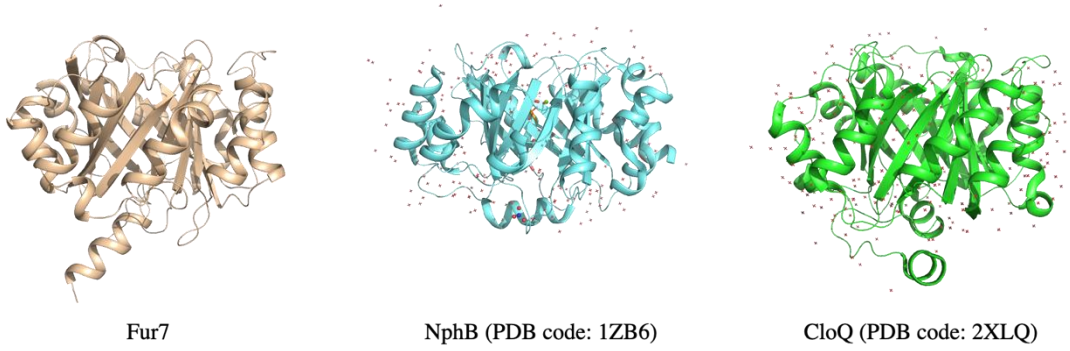


Figure 4.6 3D structures of Fur7 and its homologs. NphB is shown in aquamarine, CloQ is shown in green.

Since divalent cations do not significantly affect the enzyme activity of Fur7, and NphB uses Mg^{2+} to catalyze isoprenyl diphosphate substrates, while CloQ catalyzes reactions without metal ion binding, I conducted a detailed comparison of the active sites of Fur7, NphB, and CloQ. Upon investigating the corresponding residues of Fur7 and NphB in the active site, I discovered that the residues involved in substrate recognition in NphB are highly conserved in Fur7 (Figure 4.7). For instance, Asp62 in NphB, which coordinates to Mg^{2+}, corresponds to Asp65 in Fur7, rationalizing the enhanced activity of Fur7 with the addition of Mg^{2+}. Additionally, residues Lys119, Asn173, and Arg228 that interact with the β-phosphate of GSPP in NphB are also conserved in Fur7, with respective counterparts Lys121, Asn176, and Arg230. Furthermore, Tyr216 and Lys284, which stabilize the α-phosphate of GSPP in NphB, correspond to Tyr218 and Lys281 in Fur7, respectively. Notably, the diphosphate binding site of NphB contains fewer positively charged

residues compared to Fur7. This aspect will be further discussed in the comparison between CloQ and Fur7.

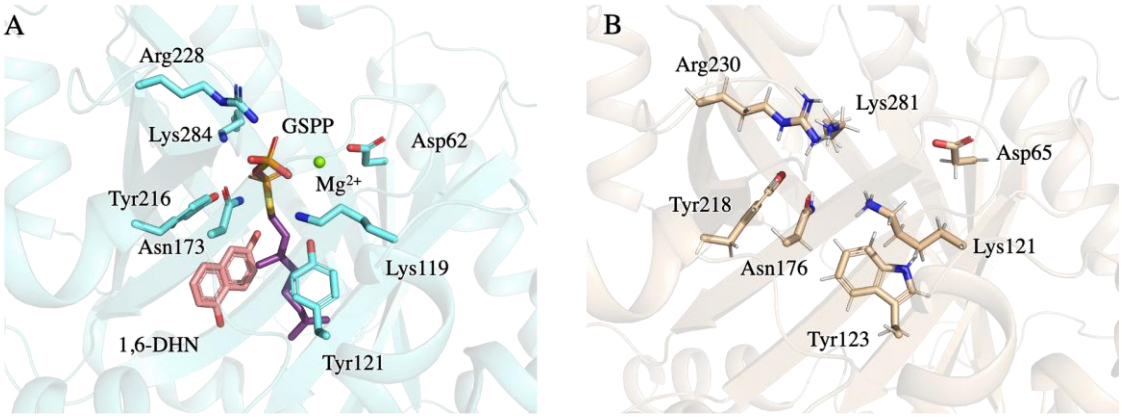


Figure 4.7 Active site structure of NphB and Fur7. NphB (A), Fur7 (B). The key residues are shown in stick model. Carbon atoms in NphB and Fur7 residues are shown in aquamarine and tints, respectively. Oxygen, nitrogen, and hydrogen atoms are shown in red, blue, and white, respectively.

As the ABBA-type superfamily exhibits a high structural similarity among its members, I will avoid redundant discussions of residues conserved in both NphB and CloQ for simplicity. CloQ is an Mg^{2+}-independent enzyme that interacts with diphosphate using positively charged residues instead of Mg^{2+}. The crystal structure shows that CloQ has more positively charged residues compared to NphB. The residues Lys54, Arg66, Lys120, Arg229, and Lys279 which play a role in stabilizing the diphosphate in CloQ, are also well conserved in Fur7, with corresponding counterparts of Lys53, Arg65, Lys121, Arg230, and Lys281. Additionally, the substitution of Ser64 in CloQ with Asp65 in Fur7 suggests the importance of Asp65 in coordinating to Mg^{2+} (Figure 4.8).

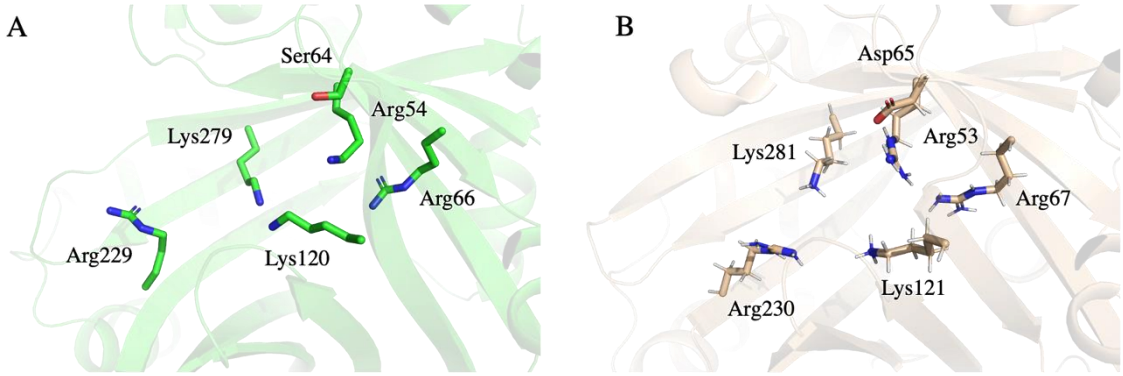
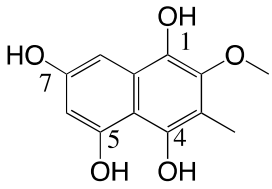


Figure 4.8 Active site structure of CloQ and Fur7. CloQ (A), Fur7 (B). The key residues are shown in stick model. Carbon atoms in CloQ and Fur7 residues are shown in green and tints, respectively. Oxygen, nitrogen, and hydrogen atoms are shown in red, blue, and white, respectively.

4.2.3 Identification of the Reactive Form of MMT

The pK_a values of all hydroxy groups of MMT (Table 4.3) calculated by using the software program Epik (Release 2021-1)[96,97] suggested that MMT possibly react in monoanionic form, and the hydroxy group at C5 was deprotonated. Therefore, MMT was prepared in C4-OH, C5-O^-.

Table 4.3 pK_a values of all hydroxy groups of MMT

Structure	Position	pKa
	1-OH	10.68
	4-OH	7.33
	5-OH	6.58
	7-OH	9.44

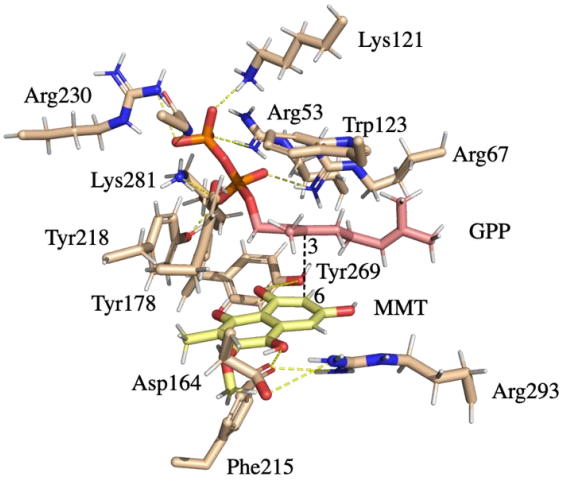


Figure 4.9 MD simulations of the Fur7 model complexed with GPP and MMT. Active site structures of Fur7-GPP-MMT (C4-OH, C5-O^-).

MD simulation revealed a stable binding mode of MMT in Fur7 (Figure 4.9). Moreover, the C-terminal α-helix of Fur7 acted like a cap enclosing the MMT and GPP in the active site (Figure 4.10 A). Additionally, the C6 atom of MMT in the complex was located at an average distance of 3.94 ± 0.43 Å from the C3 atom of GPP (Figure 4.10 B), which is sufficiently close for the prenylation reaction. Asp164 and Tyr269 form hydrogen bonds with MMT, positively charged residues Arg53, Arg67, Lys121, Arg230, and Lys281 bind the diphosphate of GPP. Additionally, the aromatic moieties of Trp123, Tyr178, Phe215, and Tyr218 create an "aromatic chamber" to stabilize the ligand in the active site. Furthermore, Asp164 and Arg293 form a hydrogen bond,

possibly acting as a gate to enclose the ligand within the active site. Thus, I conducted a QM/MM analysis using the MD simulation structure.

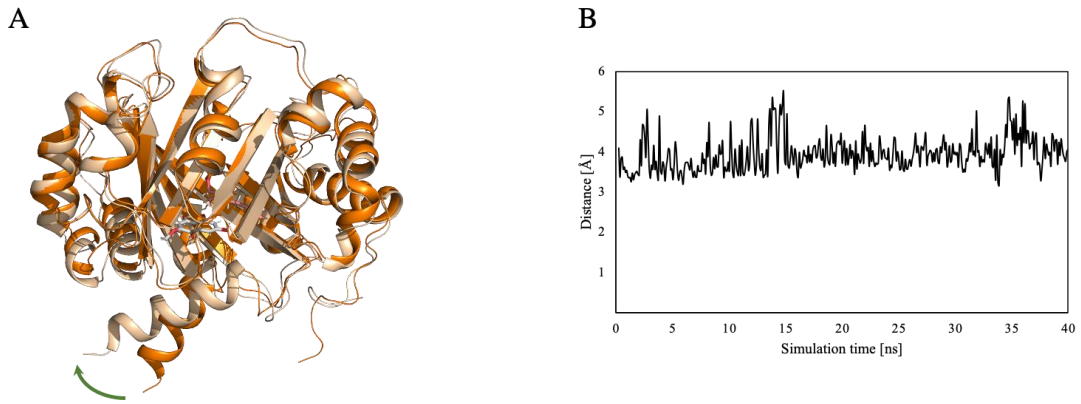


Figure 4.10 MD simulations of the Fur7 model complexed with GPP and MMT. Superpositions of Fur7 before and after MD simulation. Orange and tint are the structure before and after MD simulation, respectively. (A) Plot of the distance between the C3 of GPP and C6 of MMT, taken from the unstrained 40-ns MD simulation of the Fur7-GPP-MMT (C4-OH, C5-O^-) complex. (B)

The 2D PES scan employs the QM/MM methods (Figure 4.11). The calculated $\Delta E^{\ddagger}$ from the reactant state to the C6-intermediate was nearly 30 kcal mol^{-1}, suggesting that this reaction pathway cannot occur at room temperature. Additionally, I also detected the O5-product, with a calculated ΔE that was 7.38 kcal mol^{-1} lower than the C6-intermediate, and the calculated $\Delta E^{\ddagger}$ was also lower than that of the C6-produce reaction. This observation indicated that the reaction would produce the O5-product rather than C6-product. However, experimental analysis showed that the reaction only produced the C6-product, which contradicted the calculation results. Therefore, the reaction should not occur in this Fur7-GPP-MMT (C4-OH, C5-O^{-}) complex model. Furthermore, since the negatively charged position of anionic MMT would readily bind the geranyl group and the pK_a of 5-OH is 6.58, comparable to 7.00, I proposed that MMT reacts in its neutral form rather than in its anionic form.

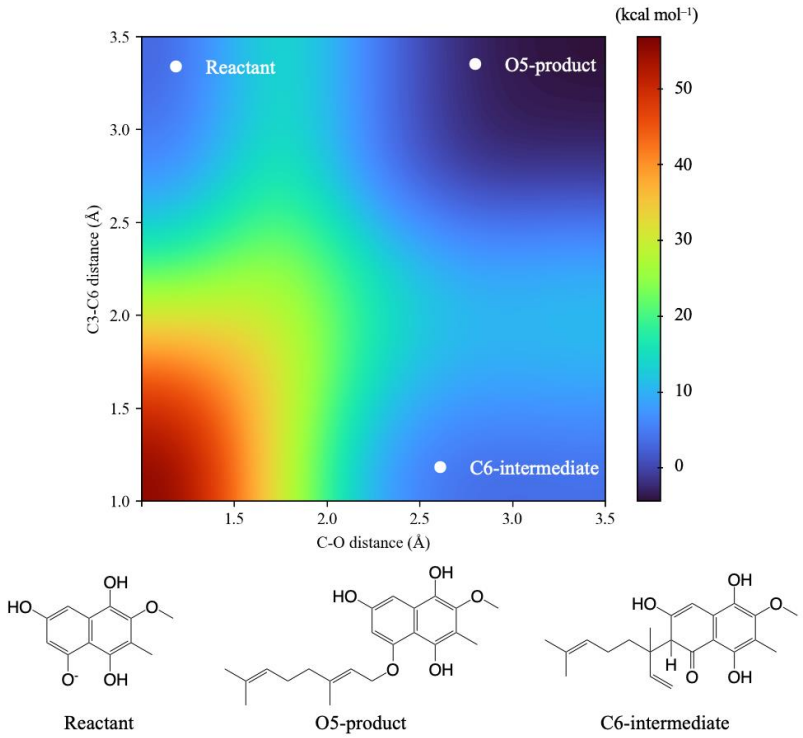


Figure 4.11 PES for the Fur7-GPP-MMT system. The QM/MM based energy profile drawn according to the two distances (O_1-C_1 and C3-C6) in the QM layer. Marked stationary points correspond to the structures shown below, energies are shown in kcal mol^{-1}.

4.2.4 MD Simulations and QM/MM Study on the Geranyl Group Transfer Step Catalyzed by the

Fur7 Complex with GPP and MMT

After identifying the reactive form of MMT, I assumed that there was no deprotonation occurring in the hydroxy group of MMT. Subsequently, a retro-biosynthetic analysis was carried out using the Fur7-OPP^{4-}-IM2 complex model. MD simulations revealed that the key residues in the active site of Fur7 exhibit a similar orientation and stereochemical structure to its homolog NphB (Figure 4.12). Positively charged residues, including Lys121, Arg230, and Lys281, bound

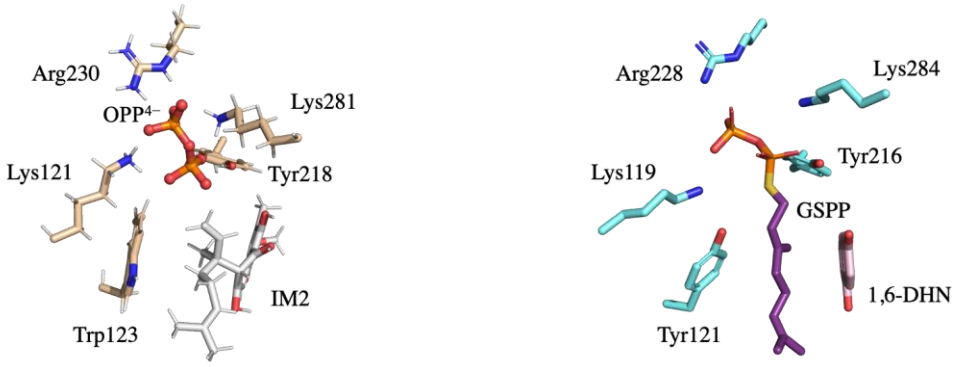


Figure 4.12 Active site structure of Fur7 and NphB. Fur7 (A), NphB (B). Carbon atoms in Fur7 and NphB residues are shown in tints and aquamarine, respectively. Oxygen, nitrogen, and hydrogen atoms are shown in red, blue, and white, respectively.

to the diphosphate, similar to Lys119, Arg228, and Lys284 in NphB. Tyr218 located near the O atom of diphosphate, corresponding to the position of Tyr216 in NphB. Notably, the aliphatic hydrocarbon chain of IM2 was confined between Trp123 and the aromatic-ring moiety of IM2, just as the hydrocarbon chain of GSPP was trapped between Tyr121 and 1,6-DHN. This "aromatic chamber" effectively protected and stabilized the carbocation during the reaction.

To further investigate the geranyl group transfer step and uncover the prenylation reaction pathway, I conducted calculations of $\Delta E^{\ddagger}$ and ΔE using the ONIOM method. For the sake of brevity, this study did not delve into the C-O bond cleavage process, as it typically applies to all prenyltransferases and it is the rate-limiting chemical step for a great number of aromatic prenyltransferases. Instead, I focused solely on the unique geranyl group transfer step in Fur7. The ONIOM calculations revealed that in the IM1 state, the geranyl group was transferred to MMT through a transition state, leading to the IM2 (Figure 4.13 A-C). In the IM1 state (Figure 4.13 A), the distance between the C6 atom of MMT and the C3 atom of the geranyl group was 2.79 Å. In TS (Figure 4.13 B), the geranyl group moved closer to the C6 atom of MMT at 2.55 Å. In the IM2 state (Figure 4.13 C), the geranyl group formed a covalent bond with the C6 atom of MMT. The calculated $\Delta E^{\ddagger}$ of TS from the IM1 state was 0.42 kcal mol^{-1} (Figure 4.13 D), and the calculated ΔE of the IM2 state was –11.42 kcal mol^{-1}. The small calculated $\Delta E^{\ddagger}$ indicated that the

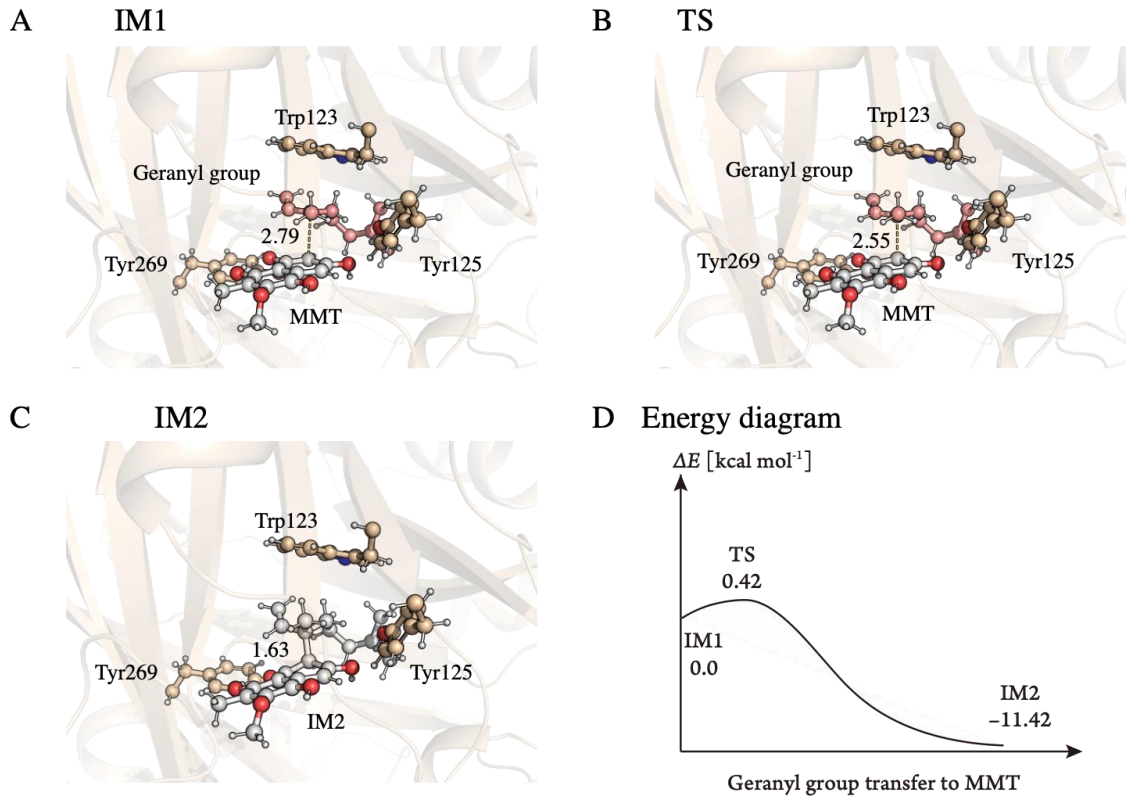


Figure 4.13 QM/MM analysis of the prenylation reaction catalyzed by Fur7. A–C) Structures of the IM1 (A), TS (B), and IM2 (C). Bond lengths are shown in angstroms. (D) Energy diagram of geranyl transfer to MMT.

prenyl transfer step can readily occur at room temperature. The calculated $\Delta E^{\ddagger}$ of Fur7 was also comparable to its homologs, NphB and CloQ, with $\Delta E^{\ddagger}$ less than 1.00 kcal mol^{-1}.[66,122] Therefore, the QM/MM results suggested that Fur7 followed an S_N1-like mechanism in a similar manner to its homologs. This mechanism involves the dissociation of a weakly stable carbocation from GPP, which is then nucleophilically captured in the prenylation step by MMT. For the extra proton transfer step in Fur7, based on the reaction mechanism of NphB and CloQ, I propose that the prenylated product is subsequently rearomatized through a water-mediated proton transfer pathway. This proposal is supported by the presence of several water molecules in the active site during MD simulations. However, I did not include them in the QM region at this time because, in the chosen snapshot, there is a lack of properly positioned water molecules. It is possible to obtain a more ideal complex model structure by increasing the number of MD simulations.

Unlike the QM/MM study for Fur6 in Sections 3.2.2 and 3.2.3, I conducted a retro-biosynthetic analysis for Fur7, as mentioned earlier. I initially attempted a QM/MM study following the biosynthetic pathway, similar to the study on Fur6. However, I encountered difficulties with the calculations, including convergence issues and the formation of unconventional structures that were not conducive to the intended reaction. During the calculation, I observed that these

difficulties arise from the high conformational flexibility of the geranyl group. In contrast, the conformations of the intermediate or product states, where the geranyl group is already bonded to MMT, are more rigid and relatively fixed. Therefore, I applied the retro-biosynthetic analysis this time. My study demonstrates that retro-biosynthetic analysis could be a more efficient approach, particularly when dealing with flexible conformations during calculations.

4.2.5 Estimation of the Reaction Pathways to Produce Normal and Reverse Products

In prenylation, the product produced by the reaction occurs at C1 secondary carbon of the allylic moiety named the "normal product," whereas the product produced by the reaction occurs at C3 tertiary carbon of the allylic moiety named the "reverse product."[85] Unlike many other aromatic prenyltransferases, Fur7 can catalyze prenylation to produce a "reverse" product rather than a "normal" one (Figure 4.14). This special catalytic property interests us in investigating the mechanism underlying the regioselectivity of Fur7.

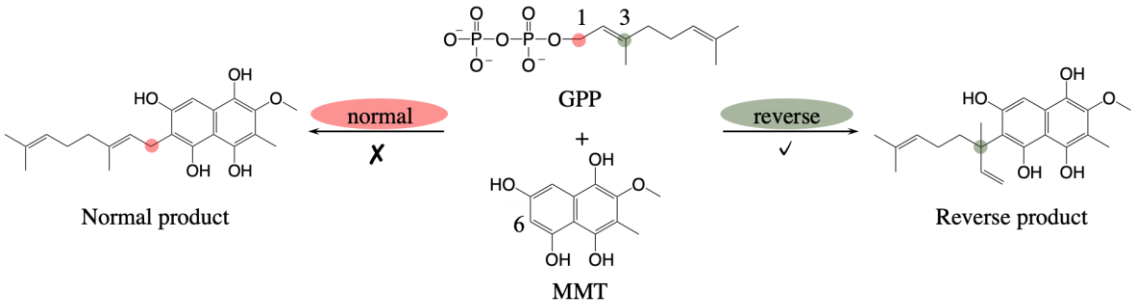


Figure 4.14 Normal and reverse prenylation with GPP and MMT

Initially, I estimated the energy barriers for reactions that produce normal and reverse products *in vacuo*, respectively. Drawing on my experience with the QM/MM analysis mentioned in Section 4.2.4, I conducted the retro-biosynthetic analysis, focusing on the prenyl transfer step. For convenience, the reactions that produce the normal and reverse products were denoted as the C1- and C3-reaction, respectively. The calculated $\Delta E^{\ddagger}$ of the TS from the C1-IM1 state was 0.06 kcal mol^{-1}, and the calculated ΔE of the C1-IM2 state was –18.51 kcal mol^{-1} (Figure 4.15). The

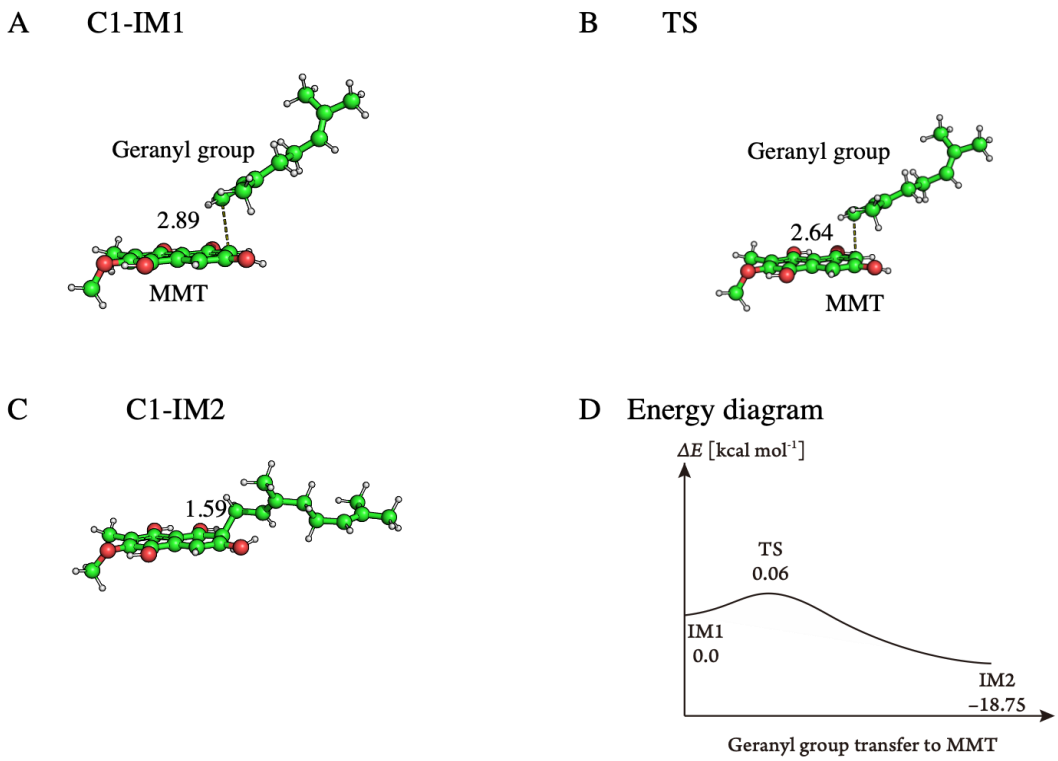


Figure 4.15 QM analysis of the prenylation reaction which produces normal product. A–C) Structures of the C1-IM1 (A), TS (B), and C1-IM2 (C). Bond lengths are shown in angstroms. (D) Energy diagram of geranyl transfer

calculated $\Delta E^{\ddagger}$ of TS from the C3-IM1 state was 4.89 kcal mol^{-1}, and the calculated ΔE of the C3-IM2 state was –7.22 kcal mol^{-1} (Figure 4.16). Calculated results indicated that the C1-reaction should be the predominant reaction rather than the C3-reaction; however, these findings were inconsistent with the experiment data. Considering the $\Delta E^{\ddagger}$ of C3-reaction calculated by the QM/MM methods was 0.42 kcal mol^{-1}, which is 4.47 kcal mol^{-1} lower than calculated *in vacuo*,

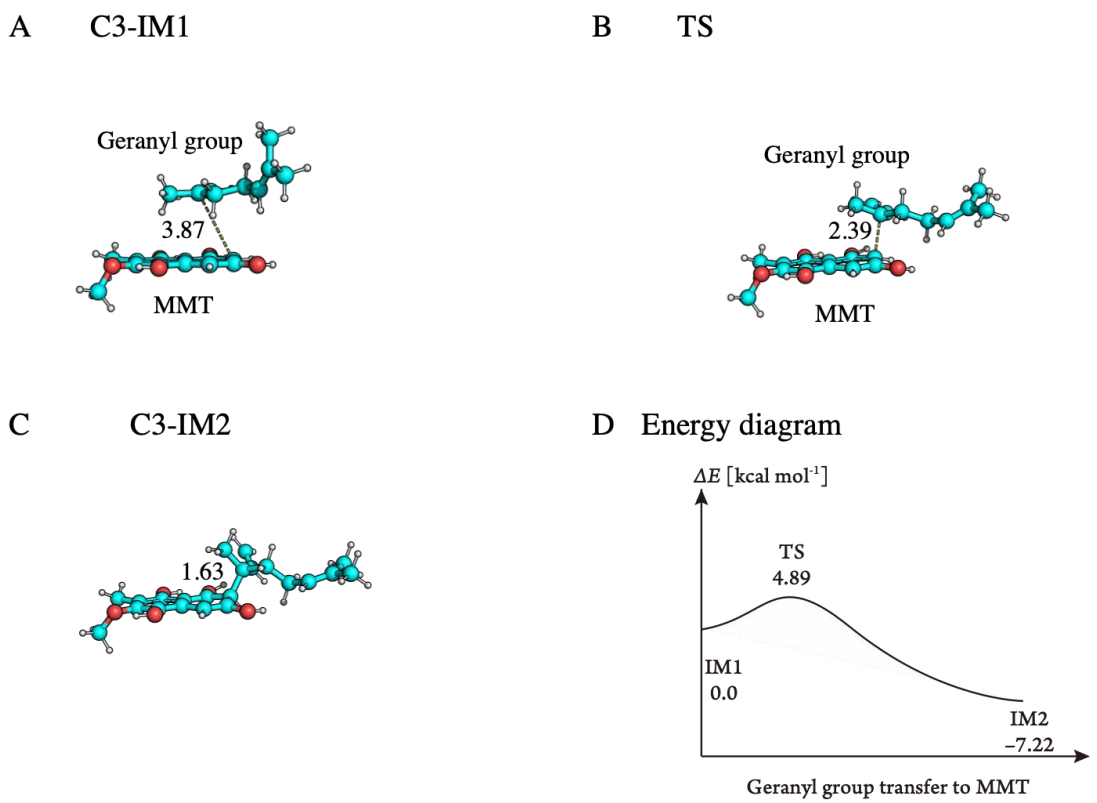


Figure 4.16 QM analysis of the prenylation reaction which produces reverse product. A–C) Structures of the C3-IM1 (A), TS (B), and C3-IM2 (C). Bond lengths are shown in angstroms. (D) Energy diagram of geranyl transfer to MMT.

indicating that the enzyme decreased the energy barrier through protein-ligand interactions. The residues in the active site could rearrange the conformation of the geranyl group to make the reaction more efficient. Moreover, compared to C3-IM1, the geranyl group in C1-IM1 was not located parallel to the aromatic ring of MMT, resulting in the loss of π–π stacking, and occupying a larger volume which cannot be accommodated in the active site (Figure 4.15).

Subsequently, MD simulations for the Fur7-GPP-MMT complex revealed that the C1 atom of GPP in the complex was located at an average distance of 4.79 ± 0.49 Å from the C6 atom of MMT, while the C3 atom of GPP in the complex was positioned at an average distance of 3.78 ± 0.37 Å from the C6 atom of MMT (Figure 4.17). The closer proximity between the C3 atom of GPP and the C6 atom of MMT could structurally explain why the reverse product is favored over the normal product, as this close distance facilitates the reaction more easily.

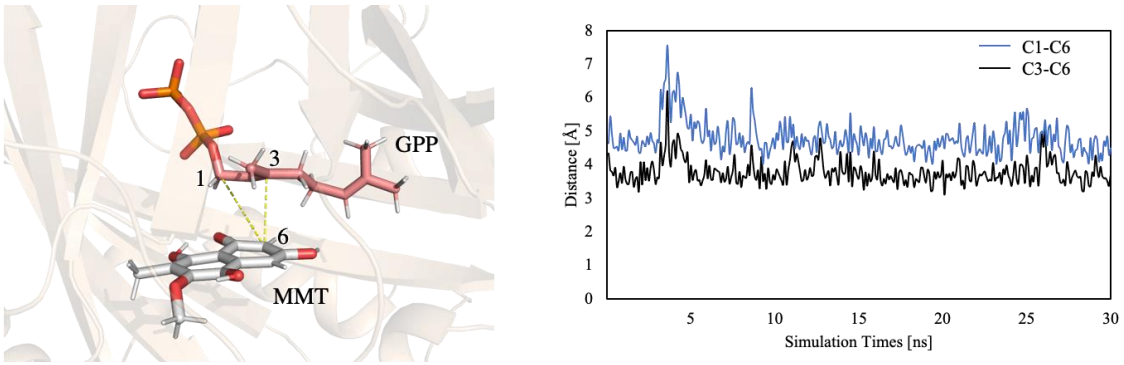


Figure 4.17 MD simulations of the Fur7 model complexed with GPP and MMT. The structure of MMT and GPP (A). Plots of the distance between the C3 of GPP and C6 of MMT, as well as between C1 of GPP and C6 of MMT, are depicted in black and blue, respectively (B).

To further investigate the complex structure in the IM1 state and confirm whether the C1 atom was still located farther than the C3 atom of GPP to the C6 atom of MMT, I conducted MD simulations using the Fur7-OPP^{4-}-Geranyl group-MMT complex model. As expected, compared to the average distance between the C1 atom of the geranyl group and the C6 atom of MMT, the C3 atom of the geranyl group was 1.29 Å closer to the C6 atom of MMT (Figure 4.18). This closer distance facilitates the production of the reverse product.

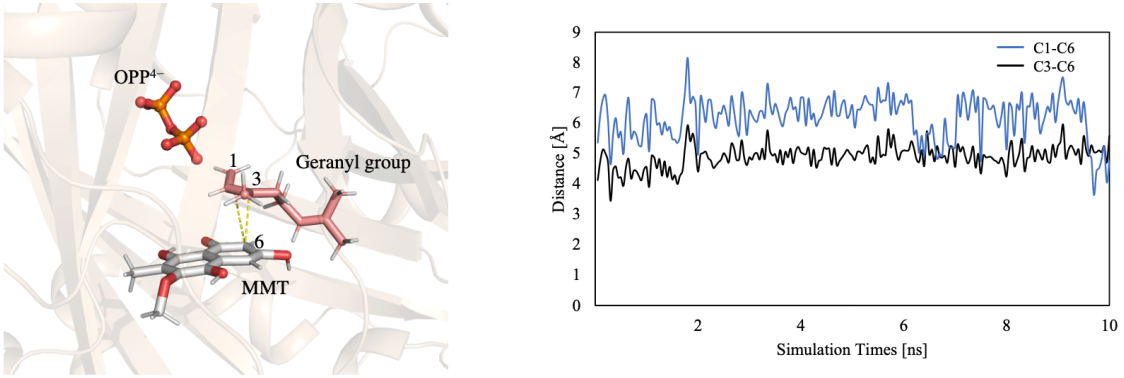


Figure 4.18 MD simulations of the Fur7 model complexed with OPP^{4-}, geranyl group, and MMT. The complex structure of IM1 (A). Plots of the distance between the C3 of GPP and C6 of MMT, as well as between C1 of GPP and C6 of MMT, are depicted in black and blue, respectively (B).

Therefore, MD simulations revealed that Fur7 can catalyze prenylation to produce a "reverse" product because the C3-GPP stays closer to C6-MMT than C1-GPP. I also suggest that further QM/MM or QM/MM MD studies could provide a more comprehensive clarification of the catalytic mechanisms.

4.3 Conclusion

In this study, I systematically explored the reaction mechanism and regioselectivity of Fur7. Firstly, through HOMO energy calculations, I clarified the original substrate of Fur7 should be MMT rather than MMF. This finding theoretically explains the experimental observation that prenylation was significantly enhanced with the addition of dithionite to MMF. Secondly, the 3D structure of Fur7 was predicted using AlphaFold2, revealing that Fur7 possesses the representative "prenyltransferase barrel" and belongs to the ABBA-type superfamily, akin to its homologs NphB and CloQ. Moreover, Fur7 features Asp65, corresponding to Asp62 in NphB, enabling Fur7 to coordinate with Mg^{2+}. Fur7 also has more positively charged residues in the active site than NphB, similar to CloQ, allowing Fur7 to react without the need for a metal ion. Thirdly, through MD simulations and QM/MM analysis, I found that the natural substrate MMT should react in the neutral form rather than the C4-OH, C5-O^- form. Positively charged residues, Lys121, Arg230, and Lys281, are crucial for diphosphate binding. Trp123 and MMT sandwich the geranyl group of GPP. Moreover, by cooperating with Tyr218 and other aromatic residues in the active site, a π chamber is formed that significantly stabilizes and protects the carbocation during the reaction. Fourthly, using the QM/MM analysis, I proposed the reaction mechanism of

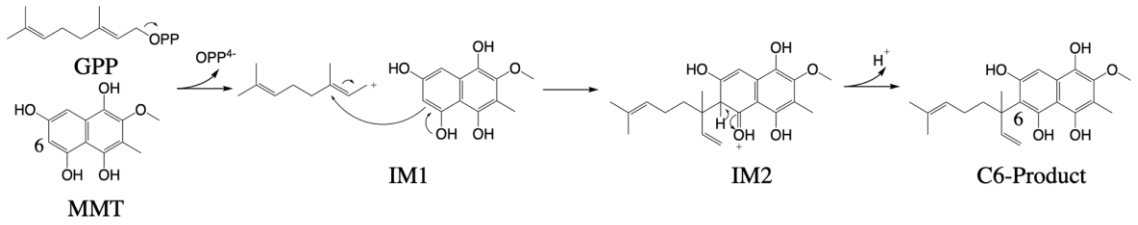


Figure 4.19 Proposed mechanism for the prenylation reaction of MMT catalyzed by Fur7.

Fur7 (Figure 4.19). The reaction follows the S_N1-like mechanism, similar to other known prenyltransferases.[66,122] This mechanism involves the dissociation of a weakly stable carbocation from GPP, which is then nucleophilically captured in the prenylation step by MMT with a $\Delta E^{\ddagger}$ of approximately 0.42 kcal mol^{-1}. Considering the homologs of Fur7 as references, the deprotonation for the extra proton may follow a water-mediated proton transfer pathway. Finally, QM calculations and MD simulations on the regioselectivity of Fur7 indicated that the residues in the active site could rearrange the conformation of the geranyl group to make the reaction more efficient. The C3 atom of GPP is closer to the C6 atom of MMT than the C1 atom of GPP, which may be one of the reasons that only the reverse product is produced. I believe that by conducting simulations with modified models or adopting alternative computational strategies in future studies, the entire mechanism and regioselectivity of Fur7 will be clarified more thoroughly.

Prenylation is a widely recognized and critical reaction in the biosynthesis of natural products, making a significant contribution to the enhancement of bioactivity and structure. A

comprehensive understanding of prenyltransferases is essential for the rational engineering of enzymes and the development of compounds that can meet therapeutic and other industrial requirements.

5 Conclusion

In conclusion, this dissertation conducted a theoretical study on the chemistry of *C*-methyltransferase Fur6 and aromatic prenyltransferase Fur7, both of which are from the biosynthetic gene cluster of furaquinocin D. The findings, as outlined in Chapters 3 and 4, show that computational strategies, such as MD simulations and QM/MM methods, were applied and successfully addressed the substrate binding modes and reaction pathways of these two enzymes. The original substrate and substrate specificity have also been analyzed, which is challenging to explain using experimental strategies in previous studies.

In Chapter 3, I focused on the reaction mechanism and substrate specificity of *C*-methyltransferase Fur6. Firstly, through docking simulation, MD simulations, and QM/MM analysis, I identified that the natural substrate PHN should react in the C4-O^-, C5-OH form rather than the neutral form. This deprotonation made Tyr137 and Trp150 form hydrogen bonds with the substrate, and with the contribution of Asp252, PHN could be stabilized in the active pocket (Figure 3.5). Furthermore, the average distance between the C3 atom of PHN and the transferrable methyl group of SAM was 3.34 ± 0.17 Å (Figure 3.6), which is sufficient for the methylation reaction to occur. Secondly, by using the QM/MM analysis and free energy calculation with the PCM method, I successfully clarified that the methylation catalyzed by Fur6 follows the S_N2 mechanism, similar to other known methyltransferases (Figure 3.10).[59,117] During the reaction, the nucleophilic attack from the C3 atom on the methyl group of SAM produces SAH and an intermediate in which the C3 atom is methylated, with a $\Delta E^{\ddagger}$ of approximately 8.6 kcal mol^{-1}. The methylated intermediate is then rearomatized automatically after it is presumably ejected into the water solvent. Thirdly, pK_a prediction and QM calculation on energy barrier indicate that, among the analogs chosen from the biosynthetic intermediates of furaquinocin D, despite 8-AF, 8-FDA, and PND forming the intermolecular interactions similar to those of PHN, they exhibit higher potential energy barriers. This leads to a decreased reaction rate compared to PHN, or they may not even be catalyzed by Fur6. Furthermore, the results of the HOMO energy of PHN and its analogs also strongly support that PHN should be the genuine substrate for Fur6. Studies on methylation have been conducted for several years, while there still remain unrevealed parts. In the past decade, the QM/MM method has been utilized to clarify the methyltransferase mechanism; however, most of them concentrated on the methylation of protein modification,[123,124] and seldom of them were focused on methylation in biosynthetic pathways,[59,125] especially the methylation for the aromatic system. This study extends the knowledge of the QM/MM study of *C*-methyltransferase in the aromatic system. Moreover, adding a methyl group to an aromatic system in organic or industrial synthesis has been a challenge for a long time. I believe that with

the catalytic mechanism of more *C*-methyltransferases like Fur6 being clarified, enzyme engineering could be a potential tool to address this problem.

In Chapter 4, I investigate the structural features, reaction mechanism, and regioselectivity of aromatic prenyltransferase Fur7. Firstly, through HOMO energy calculations, I determined that the authentic substrate for Fur7 is MMT, not MMF. This theoretical clarification aligns with experimental observations, in which the prenylation was enhanced when dithionite was added to MMF, giving insights into the observed phenomena. Secondly, the 3D structure of Fur7 was predicted using AlphaFold2, revealing that Fur7 possesses the representative "prenyltransferase barrel" and belongs to the ABBA-type superfamily (Figure 4.5). Moreover, the model structure shows that Fur7 possesses a great number of positively charged residues in the active site akin to its homolog CloQ and reserves the Asp62 which can help coordinate with Mg^{2+}. That could explain why Fur7 exhibits both the features of Mg^{2+}-dependent and Mg^{2+}-independent aromatic prenyltransferase. Thirdly, through docking simulation and MD simulations, I found that the natural substrate MMT should react in the neutral form rather than the C4-OH, C5-O^- form. Positively charged residues, Lys121, Arg230, and Lys281, function to bind the diphosphate of GPP. Trp123 and MMT sandwich the geranyl group of GPP. Moreover, a π chamber formed by Trp123, Tyr218, and other aromatic residues in the active site significantly stabilizes and protects the carbocation during the reaction (Figure 4.9). Fourthly, through MD simulations and QM/MM method used in retro-biosynthetic analysis, I propose that the reaction catalyzed by Fur7 follows an S_N1-like mechanism, similar to other known prenyltransferases.[66,122] The mechanism involves the dissociation of a carbocation from GPP, which is then nucleophilically captured in the prenylation step by MMT with a $\Delta E^{\ddagger}$ of approximately 0.42 kcal mol^{-1}. Moreover, the deprotonation for the extra proton may follow a water-mediated proton transfer pathway. Finally, MD simulations on the regioselectivity of Fur7 indicated that the C3 atom of GPP is closer to the C6 atom of MMT than the C1 atom of GPP, which may be one of the reasons that only the reverse product is produced. In the past several decades, research on prenyltransferase has significantly advanced, unraveling their diverse roles and implications. Multiple research strategies such as experimental and computational tools have been used to elucidate the chemistry behind the prenyltransferase; however, the complex reaction mechanism and complicated substrate specificity make them have not been studied sufficiently.[31,122] The study of Fur7 enriches research on the prenyltransferase mechanism and broadens the understanding of the chemistry of aromatic prenyltransferases. Although some mysteries behind the mechanism and regioselectivity have been unveiled, this study still provides a reference for subsequent computational research, such as the choice of methods and the conduct of systems.

In this study, I primarily described the protein-ligand interaction, reaction mechanism, and reaction specificity of *C*-methyltransferase Fur6 and aromatic prenyltransferase Fur7. My study

highlights the potential of computational strategies in enzymatic research, as the substrates of Fur6 and Fur7 are prone to oxidation, making the reaction challenging to analyze through experimental methods. Moreover, simulation studies, such as MD simulations, provide insights into dynamic molecular processes, revealing structural changes and interactions over time. QM/MM analysis enables a detailed examination of reaction mechanisms at the quantum level, accounting for the electronic behavior of active sites. These methods enhance the understanding of enzymatic function, substrate binding, and reaction pathways with atomistic precision. Additionally, computational approaches facilitate the exploration of complex systems that may be challenging or impossible to investigate solely through experimental tools, contributing to advancements in enzyme engineering, efficient drug discovery, and other fields.

6 Appendix

Appendix 1.

Table 6.1 Gaussian input structures for *in vacuo* calculation

(1) 8-AF and SAM-1

C	-0.03200000	0.19300000	0.37600000
C	-1.13300000	-0.37700000	-0.29800000
C	-2.39100000	-0.31100000	0.27400000
H	-1.02100000	-0.82000000	-1.28200000
C	-0.17400000	0.84600000	1.58000000
C	-1.45700000	0.90700000	2.20400000
C	-2.58800000	0.28800000	1.55700000
C	-3.88000000	0.28400000	2.19400000
C	-3.98500000	0.91100000	3.43100000
H	-4.95800000	1.02600000	3.89900000
C	-2.87200000	1.48100000	4.09500000
C	-1.62100000	1.52100000	3.48100000
C	-3.53200000	-2.67000000	7.32500000
H	-3.72800000	-2.20100000	8.29000000
H	-2.84500000	-2.04800000	6.75000000
S	-5.07300000	-2.70800000	6.34900000
C	-4.86400000	-1.17200000	5.35100000
H	-5.82300000	-0.99500000	4.86100000
H	-4.09900000	-1.36600000	4.60000000
H	-4.56100000	-0.34400000	5.99700000
C	-6.35800000	-2.16900000	7.52100000
H	-7.21400000	-1.88700000	6.90100000
H	-5.99900000	-1.29700000	8.06800000
H	-6.63638495	-2.95822134	8.18773196
H	-3.08662182	-3.63475335	7.45065528
N	0.99353701	1.46035009	2.22832963
H	1.69360774	0.76341028	2.38381552
H	1.36309597	2.17822768	1.63834832
O	-0.60766538	2.01852169	4.03705366
O	-4.87125551	-0.26649086	1.64815855
O	-3.50310039	-0.85441269	-0.44212525
H	-3.20470555	-1.19040334	-1.29045938
O	1.26480289	0.10360075	-0.21999499
O	-3.10709346	1.93334501	5.43104455
H	-3.81751366	2.57902508	5.43041939

(2) FLA and SAM-1

O	1.23000000	0.10600000	-0.20400000
H	1.22900000	-0.69100000	-0.75600000
C	-0.03200000	0.19300000	0.37600000
C	-1.13300000	-0.37700000	-0.29800000
C	-2.39100000	-0.31100000	0.27400000
O	-3.44700000	-0.82700000	-0.40600000
H	-4.24000000	-0.64200000	0.17700000
H	-1.02100000	-0.82000000	-1.28200000
C	-0.17400000	0.84600000	1.58000000
H	0.68500000	1.29800000	2.05700000
C	-1.45700000	0.90700000	2.20400000
C	-2.58800000	0.28800000	1.55700000
C	-3.88000000	0.28400000	2.19400000
C	-3.98500000	0.91100000	3.43100000
H	-4.95800000	1.02600000	3.89900000
C	-2.87200000	1.48100000	4.09500000
C	-1.62100000	1.52100000	3.48100000
C	-3.53200000	-2.67000000	7.32500000
H	-3.72800000	-2.20100000	8.29000000
H	-2.84500000	-2.04800000	6.75000000
S	-5.07300000	-2.70800000	6.34900000
C	-4.86400000	-1.17200000	5.35100000
H	-5.82300000	-0.99500000	4.86100000
H	-4.09900000	-1.36600000	4.60000000
H	-4.56100000	-0.34400000	5.99700000
C	-6.35800000	-2.16900000	7.52100000
H	-7.21400000	-1.88700000	6.90100000
H	-5.99900000	-1.29700000	8.06800000
H	-6.63638495	-2.95822134	8.18773196
H	-3.08662182	-3.63475335	7.45065528
O	-3.10709346	1.93334501	5.43104455
O	-0.60766538	2.01852169	4.03705366
O	-4.87125551	-0.26649086	1.64815855

(3) MOP and SAM-1

① C3 reaction

C	-0.03200000	0.19300000	0.37600000
C	-1.13300000	-0.37700000	-0.29800000
C	-2.39100000	-0.31100000	0.27400000
H	-1.02100000	-0.82000000	-1.28200000
C	-0.17400000	0.84600000	1.58000000
C	-1.45700000	0.90700000	2.20400000
C	-2.58800000	0.28800000	1.55700000
C	-3.88000000	0.28400000	2.19400000
C	-3.98500000	0.91100000	3.43100000
H	-4.95800000	1.02600000	3.89900000
C	-2.87200000	1.48100000	4.09500000
O	-3.09600000	1.91200000	5.36800000
C	-1.62100000	1.52100000	3.48100000
C	-3.53200000	-2.67000000	7.32500000
H	-3.72800000	-2.20100000	8.29000000
H	-2.84500000	-2.04800000	6.75000000
S	-5.07300000	-2.70800000	6.34900000
C	-4.86400000	-1.17200000	5.35100000
H	-5.82300000	-0.99500000	4.86100000
H	-4.09900000	-1.36600000	4.60000000
H	-4.56100000	-0.34400000	5.99700000
C	-6.35800000	-2.16900000	7.52100000
H	-7.21400000	-1.88700000	6.90100000
H	-5.99900000	-1.29700000	8.06800000
H	-6.63638495	-2.95822134	8.18773196
H	-3.08662182	-3.63475335	7.45065528
O	0.96176729	1.44363308	2.21068801
H	1.45921130	1.94842289	1.56312589
O	1.26480289	0.10360075	-0.21999499
H	1.19701893	-0.33683832	-1.07030009
O	-3.50310039	-0.85441269	-0.44212525
H	-3.20266829	-1.20134404	-1.28532107
O	-0.60766538	2.01852169	4.03705366
O	-4.87125551	-0.26649086	1.64815855

② C6 reaction

C	-0.03200000	0.19300000	0.37600000
C	-1.13300000	-0.37700000	-0.29800000
C	-2.39100000	-0.31100000	0.27400000
H	-1.02100000	-0.82000000	-1.28200000
C	-0.17400000	0.84600000	1.58000000
C	-1.45700000	0.90700000	2.20400000
C	-2.58800000	0.28800000	1.55700000
C	-3.88000000	0.28400000	2.19400000
C	-3.98500000	0.91100000	3.43100000
H	-4.95800000	1.02600000	3.89900000
C	-2.87200000	1.48100000	4.09500000
C	-1.62100000	1.52100000	3.48100000
C	-3.53200000	-2.67000000	7.32500000
H	-3.72800000	-2.20100000	8.29000000
H	-2.84500000	-2.04800000	6.75000000
S	-5.07300000	-2.70800000	6.34900000
C	-4.86400000	-1.17200000	5.35100000
H	-5.82300000	-0.99500000	4.86100000
H	-4.09900000	-1.36600000	4.60000000
H	-4.56100000	-0.34400000	5.99700000
C	-6.35800000	-2.16900000	7.52100000
H	-7.21400000	-1.88700000	6.90100000
H	-5.99900000	-1.29700000	8.06800000
H	-6.63638495	-2.95822134	8.18773196
H	-3.08662182	-3.63475335	7.45065528
O	1.26480289	0.10360075	-0.21999499
O	-0.60766538	2.01852169	4.03705366
H	-0.84957626	2.84278157	4.46562661
O	-3.10709346	1.93334501	5.43104455
H	-3.84697325	2.54501924	5.43674898
O	-4.87125551	-0.26649086	1.64815855
H	-5.64380935	-0.16371795	2.20869248
O	0.96176729	1.44363308	2.21068801
O	-3.50310039	-0.85441269	-0.44212525

(4) 8-DFA and SAM-1

C	-0.17400000	0.84600000	1.58000000
C	-3.53200000	-2.67000000	7.32500000
H	-3.72800000	-2.20100000	8.29000000
H	-2.84500000	-2.04800000	6.75000000
S	-5.07300000	-2.70800000	6.34900000
C	-4.86400000	-1.17200000	5.35100000
H	-5.82300000	-0.99500000	4.86100000
H	-4.09900000	-1.36600000	4.60000000
H	-4.56100000	-0.34400000	5.99700000
C	-6.35800000	-2.16900000	7.52100000
H	-7.21400000	-1.88700000	6.90100000
H	-5.99900000	-1.29700000	8.06800000
H	-6.63638495	-2.95822134	8.18773196
H	-3.08662182	-3.63475335	7.45065528
N	0.99353701	1.46035009	2.22832963
N	1.79180485	1.89126593	1.62042833
C	-0.03200000	0.19300000	0.37600000
C	-1.13300000	-0.37700000	-0.29800000
H	-1.03564504	-0.80823017	-1.27240369
C	-2.39100000	-0.31100000	0.27400000
C	-1.62100000	1.52100000	3.48100000
C	-2.87200000	1.48100000	4.09500000
C	-3.98500000	0.91100000	3.43100000
H	-4.97099450	0.99283914	3.83845209
C	-3.88000000	0.28400000	2.19400000
C	-2.58800000	0.28800000	1.55700000
C	-1.45700000	0.90700000	2.20400000
O	1.26480289	0.10360075	-0.21999499
O	-3.50310039	-0.85441269	-0.44212525
H	-3.19912281	-1.22173114	-1.27535678
O	-3.10709346	1.93334501	5.43104455
H	-3.91682415	2.44837237	5.45717962
O	-0.60766538	2.01852169	4.03705366
O	-4.87125551	-0.26649086	1.64815855

(5) PND and SAM-1

O	1.23000000	0.10600000	-0.20400000
C	-0.03200000	0.19300000	0.37600000
C	-1.13300000	-0.37700000	-0.29800000
C	-2.39100000	-0.31100000	0.27400000
O	-3.44700000	-0.82700000	-0.40600000
H	-4.24000000	-0.64200000	0.17700000
H	-1.02100000	-0.82000000	-1.28200000
C	-0.17400000	0.84600000	1.58000000
C	-1.45700000	0.90700000	2.20400000
C	-2.58800000	0.28800000	1.55700000
C	-3.88000000	0.28400000	2.19400000
C	-3.98500000	0.91100000	3.43100000
H	-4.95800000	1.02600000	3.89900000
C	-2.87200000	1.48100000	4.09500000
O	-3.09600000	1.91200000	5.36800000
H	-2.50900000	2.68600000	5.63100000
C	-1.62100000	1.52100000	3.48100000
O	-0.51300000	2.06500000	4.08900000
H	-0.78700000	2.78300000	4.71400000
C	-3.53200000	-2.67000000	7.32500000
H	-3.72800000	-2.20100000	8.29000000
H	-2.84500000	-2.04800000	6.75000000
S	-5.07300000	-2.70800000	6.34900000
C	-4.86400000	-1.17200000	5.35100000
H	-5.82300000	-0.99500000	4.86100000
H	-4.09900000	-1.36600000	4.60000000
H	-4.56100000	-0.34400000	5.99700000
C	-6.35800000	-2.16900000	7.52100000
H	-7.21400000	-1.88700000	6.90100000
H	-5.99900000	-1.29700000	8.06800000
H	-6.63638495	-2.95822134	8.18773196
H	-3.08662182	-3.63475335	7.45065528
N	0.99353701	1.46035009	2.22832963
O	-5.00642671	-0.34155780	1.57372563
H	-5.77439905	-0.25608385	2.14338579
N	1.79180485	1.89126593	1.62042833

Appendix 2. Parameters used in MD simulations

(1) Parameters of SAM (sam.prep)

```
    0    0    2

This is a remark line
molecule.res
SAM   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  O     o     M    3   2   1     1.540   111.208  -180.000 -0.670190 8
   5  C     c     M    4   3   2     1.247   153.998    36.598  0.688060 6
   6  OXT   o     E    5   4   3     1.259   130.271    72.685 -0.670190 8
   7  CA    c3    M    5   4   3     1.584   116.486  -108.671  0.243570 6
   8  N     n4    3    7   5   4     1.507   103.335  -171.111 -0.638970 7
   9  HN1   hn    E    8   7   5     1.080    97.825   -15.412  0.361960 1
  10  HN2   hn    E    8   7   5     1.042   114.838    94.284  0.361960 1
  11  H7    hn    E    8   7   5     1.020   114.779  -136.161  0.361960 1
  12  H17   hx    E    7   5   4     1.094   108.621    75.646  0.057270 1
  13  CB    c3    M    7   5   4     1.542   113.399   -47.818 -0.256850 6
  14  HB2   hc    E   13   7   5     1.093   109.191   -51.055  0.126420 1
  15  HB3   hc    E   13   7   5     1.095   110.463  -168.289  0.126420 1
  16  CG    c3    M   13   7   5     1.547   110.723    70.846 -0.054380 6
  17  HG1   h1    E   16  13   7     1.091   111.193   -16.831  0.111310 1
  18  HG2   h1    E   16  13   7     1.093   111.956   107.465  0.111310 1
  19  SD    s4    M   16  13   7     1.859   114.025  -135.986  0.243400 16
  20  CS1   c3    3   19  16  13     1.824    99.445   179.694 -0.462690 6
  21  H14   h1    E   20  19  16     1.091   110.538   -55.113  0.207050 1
  22  H15   h1    E   20  19  16     1.092   107.347  -176.267  0.207050 1
  23  H16   h1    E   20  19  16     1.092   107.721    66.178  0.207050 1
  24  C5'   c3    M   19  16  13     1.861   104.085    75.442  0.124110 6
  25  H5    h1    E   24  19  16     1.093   106.162   125.424  0.079380 1
  26  H6    h1    E   24  19  16     1.097   105.464     8.584  0.079380 1
  27  C4'   c3    M   24  19  16     1.528   111.350  -113.671  0.043230 6
  28  C3'   c3    3   27  24  19     1.541   113.521   153.122  0.164130 6
  29  O3'   oh    S   28  27  24     1.416   111.037    84.368 -0.699130 8
  30  H9    ho    E   29  28  27     0.966   110.491  -140.353  0.470980 1
  31  C2'   c3    B   28  27  24     1.555   101.896  -155.974  0.211840 6
  32  O2'   oh    S   31  28  27     1.411   110.855   -87.403 -0.707410 8
  33  H10   ho    E   32  31  28     0.971   108.121   -39.746  0.482340 1
  34  H3    h1    E   31  28  27     1.098   112.111   147.801  0.034110 1
  35  H2    h1    E   28  27  24     1.097   109.424   -39.002  0.070060 1
  36  H1    h1    E   27  24  19     1.098   111.107   -83.170  0.155040 1
  37  O4'   os    M   27  24  19     1.426   108.583    38.548 -0.602030 8
  38  C1'   c3    M   37  27  24     1.452   109.377   157.212  0.561200 6
  39  H4    h2    E   38  37  27     1.093   109.958    99.750  0.102270 1
  40  N9    na    M   38  37  27     1.444   106.647  -144.791 -0.428120 7
```

```
  41  C4    ca    M   40  38  37     1.393   123.631   -92.270  0.620510 6
  42  N3    nb    M   41  40  38     1.333   127.628   -17.251 -0.756880 7
  43  C2    ca    M   42  41  40     1.340   111.561   179.807  0.575440 6
  44  HN3   h5    E   43  42  41     1.086   116.244  -179.659  0.097010 1
  45  N1    nb    M   43  42  41     1.343   128.039     0.933 -0.776430 7
  46  C6    ca    M   45  43  42     1.343   119.149    -0.852  0.838950 6
  47  N6    nh    B   46  45  43     1.360   118.618  -179.266 -1.005600 7
  48  H21   hn    E   47  46  45     1.011   116.290   -11.512  0.443600 1
  49  H22   hn    E   47  46  45     1.009   118.857  -159.734  0.443600 1
  50  C5    ca    M   46  45  43     1.413   118.245    -1.053 -0.139630 6
  51  N7    nc    M   50  46  45     1.392   133.495  -177.480 -0.612060 7
  52  C8    cd    M   51  50  46     1.318   104.636  -178.499  0.365010 6
  53  H11   h5    E   52  51  50     1.083   124.043  -179.491  0.103580 1


LOOP
  C1'  C2'
   C8   N9
   C5   C4

IMPROPER
   CA    O    C  OXT
  C1'   C4   N9   C8
   C5   N9   C4   N3
  HN3   N3   C2   N1
   C5   N1   C6   N6
   C6  H21   N6  H22
   C4   C6   C5   N7
  H11   N9   C8   N7

DONE
STOP
```

(2) Parameters of PHN (4-OH, 5-O^-) (phn2.mol2)

```
@<TRIPOS>MOLECULE
PHN
   22    23     1     0     0
SMALL
resp


@<TRIPOS>ATOM
  1 O1          -1.5860    -2.5800     0.0280 oh         1 PHN      -0.57371
  2 O2          -0.7450     2.9240    -0.0100 oh         1 PHN      -0.67105
  3 O3           1.7320     2.5470     0.0470 o          1 PHN      -0.78223
  4 O4          -3.6940    -0.8600     0.0070 oh         1 PHN      -0.59973
  5 O5           3.4570    -1.9180    -0.0360 oh         1 PHN      -0.67762
  6 C1           0.1990     0.7100    -0.0060 ca         1 PHN      -0.42883
  7 C2          -0.0080    -0.7060    -0.0290 ca         1 PHN      0.02922
  8 C3          -1.3420    -1.2070    -0.0100 ca         1 PHN      0.16274
  9 C4          -0.9270     1.5870    -0.0190 ca         1 PHN      0.54390
```

```
 10 C5           1.5470     1.2710     0.0260 ca         1 PHN      0.76500
 11 C6           1.1120    -1.5810    -0.0440 ca         1 PHN      -0.51552
 12 C7          -2.4140    -0.3340    -0.0160 ca         1 PHN      0.33960
 13 C8          -2.2150     1.0630    -0.0430 ca         1 PHN      -0.67569
 14 C9           2.3870    -1.0340    -0.0100 ca         1 PHN      0.57360
 15 C10          2.6190     0.3450     0.0380 ca         1 PHN      -0.78489
 16 H1           1.0020    -2.6540    -0.1700 ha         1 PHN      0.18904
 17 H2          -3.0660     1.7420    -0.0580 ha         1 PHN      0.22419
 18 H3           3.6350     0.7400     0.0710 ha         1 PHN      0.19362
 19 H4          -0.8670    -2.9580     0.5590 ho         1 PHN      0.37073
 20 H5           0.2820     3.0240     0.0130 ho         1 PHN      0.46373
 21 H6          -4.2980    -0.1020     0.0240 ho         1 PHN      0.41555
 22 H7           4.2600    -1.3750    -0.0510 ho         1 PHN      0.43835
@<TRIPOS>BOND
 1     1     8 1
 2     1    19 1
 3     2     9 1
 4     2    20 1
 5     3    10 1
 6     4    12 1
 7     4    21 1
 8     5    14 1
 9     5    22 1
10     6     7 ar
11     6     9 ar
12     6    10 ar
13     7     8 ar
14     7    11 ar
15     8    12 ar
16     9    13 ar
17    10    15 ar
18    11    14 ar
19    11    16 1
20    12    13 ar
21    13    17 1
22    14    15 ar
23    15    18 1
@<TRIPOS>SUBSTRUCTURE
 1 PHN          1 TEMP              0 ****  ****    0 ROOT
```

(3) Parameters of PHN (4-O^-, 5-OH) (phn.mol2)

```
@<TRIPOS>MOLECULE
PHN
   22    23     1     0     0
SMALL
resp


@<TRIPOS>ATOM
  1 O1          -1.6180    -2.5270    -0.0000 oh         1 PHN      -0.66987
  2 O2          -0.7510     2.9830     0.0000 o          1 PHN      -0.72079
```

```
  3 O3           1.8810     2.4960    -0.0000 oh         1 PHN      -0.61215
  4 O4          -3.6950    -0.8030    -0.0000 oh         1 PHN      -0.69054
  5 O5           3.3750    -2.0680     0.0000 oh         1 PHN      -0.69103
  6 C1           0.2080     0.7560    -0.0000 ca         1 PHN      -0.37475
  7 C2          -0.0410    -0.6740    -0.0000 ca         1 PHN      0.01116
  8 C3          -1.3660    -1.1470    -0.0010 ca         1 PHN      0.09965
  9 C4          -0.9120     1.7370     0.0000 ca         1 PHN      0.72898
 10 C5           1.5670     1.1590    -0.0000 ca         1 PHN      0.59366
 11 C6           1.0430    -1.6040    -0.0000 ca         1 PHN      -0.37598
 12 C7          -2.4150    -0.2370    -0.0000 ca         1 PHN      0.39701
 13 C8          -2.2210     1.1380    -0.0000 ca         1 PHN      -0.81917
 14 C9           2.3370    -1.1460     0.0000 ca         1 PHN      0.52114
 15 C10          2.6150     0.2350     0.0000 ca         1 PHN      -0.75565
 16 H1           0.8410    -2.6690     0.0000 ha         1 PHN      0.20982
 17 H2          -3.0730     1.8180    -0.0010 ha         1 PHN      0.20384
 18 H3           3.6490     0.5880    -0.0000 ha         1 PHN      0.17434
 19 H4          -2.5850    -2.6010     0.0030 ho         1 PHN      0.44991
 20 H5           2.8510     2.5460     0.0010 ho         1 PHN      0.41138
 21 H6          -4.3220    -0.0660     0.0070 ho         1 PHN      0.45936
 22 H7           4.2020    -1.5630     0.0010 ho         1 PHN      0.44968
@<TRIPOS>BOND
 1     1     8 1
 2     1    19 1
 3     2     9 1
 4     3    10 1
 5     3    20 1
 6     4    12 1
 7     4    21 1
 8     5    14 1
 9     5    22 1
10     6     7 ar
11     6     9 ar
12     6    10 ar
13     7     8 ar
14     7    11 ar
15     8    12 ar
16     9    13 ar
17    10    15 ar
18    11    14 ar
19    11    16 1
20    12    13 ar
21    13    17 1
22    14    15 ar
23    15    18 1
@<TRIPOS>SUBSTRUCTURE
 1 PHN          1 TEMP              0 ****  ****    0 ROOT
```

(4) Parameters of MMT (5-O^-) (MMT3.prep)

```
   0    0    2

This is a remark line
```

```
molecule.res
MMT   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  C12   c3    M    3   2   1     1.540   111.208  -180.000 -0.031710 6
   5  H6    h1    E    4   3   2     1.096   149.043    18.724  0.068680 1
   6  H7    h1    E    4   3   2     1.099    53.011   -44.861  0.068680 1
   7  H8    h1    E    4   3   2     1.098   101.184  -149.793  0.068680 1
   8  O1    os    M    4   3   2     1.425    65.893   101.621 -0.282420 8
   9  C5    ca    M    8   4   3     1.411   112.412    30.594 -0.145040 6
  10  C6    ca    S    9   8   4     1.375   116.787   -92.943  0.337270 6
  11  O4    oh    S   10   9   8     1.389   119.643     0.086 -0.609350 8
  12  H10   ho    E   11  10   9     0.974   104.065   -12.169  0.400180 1
  13  C3    ca    M    9   8   4     1.414   120.946    87.959 -0.051210 6
  14  C11   c3    3   13   9   8     1.510   120.808     2.903 -0.191050 6
  15  H3    hc    E   14  13   9     1.104   113.679  -113.636  0.053120 1
  16  H4    hc    E   14  13   9     1.103   112.357   125.110  0.053120 1
  17  H5    hc    E   14  13   9     1.094   110.084     6.515  0.053120 1
  18  C4    ca    M   13   9   8     1.401   117.317  -179.381  0.213700 6
  19  O5    oh    S   18  13   9     1.377   118.406   179.965 -0.521880 8
  20  H11   ho    E   19  18  13     0.969   106.829    -3.077  0.387860 1
  21  C1    ca    M   18  13   9     1.423   122.739    -0.514 -0.268630 6
  22  C2    ca    M   21  18  13     1.444   118.240    -0.927 -0.001580 6
  23  C8    ca    M   22  21  18     1.410   121.509  -179.229 -0.597360 6
  24  H1    ha    E   23  22  21     1.087   119.864   179.898  0.175520 1
  25  C9    ca    M   23  22  21     1.395   118.578    -0.202  0.584800 6
  26  O3    oh    S   25  23  22     1.391   119.792   179.983 -0.639750 8
  27  H9    ho    E   26  25  23     0.969   106.246    -0.053  0.399200 1
  28  C10   ca    M   25  23  22     1.384   122.548    -0.003 -0.782480 6
  29  H2    ha    E   28  25  23     1.087   120.025  -179.923  0.219510 1
  30  C7    ca    M   28  25  23     1.440   122.892     0.084  0.782010 6
  31  O2    o     M   30  28  25     1.253   121.706  -179.901 -0.742990 8


LOOP
   C2   C6
   C7   C1

IMPROPER
   C3   C6   C5   O1
   C2   C5   C6   O4
  C11   C4   C3   C5
   C3   C1   C4   O5
   C4   C2   C1   C7
   C1   C8   C2   C6
   C2   C9   C8   H1
   C8  C10   C9   O3
   C9   C7  C10   H2
   C1  C10   C7   O2
```

```
DONE
```

(5) Parameters of GPP (gpp.prep)

```
    0    0    2

This is a remark line
molecule.res
GPP   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  O1B   o     M    3   2   1     1.540   111.208  -180.000 -0.938530 8
   5  PB    p5    M    4   3   2     1.555    55.497  -110.697  1.347940 15
   6  O2B   o     E    5   4   3     1.554   112.228   153.520 -0.938530 8
   7  O3B   o     E    5   4   3     1.536   115.147   -71.668 -0.938530 8
   8  O3A   os    M    5   4   3     1.754   104.908    40.060 -0.524750 8
   9  PA    p5    M    8   5   4     1.607   147.115    62.830  1.167160 15
  10  O1A   o     E    9   8   5     1.530   110.870   105.961 -0.865190 8
  11  O2A   o     E    9   8   5     1.522   111.404  -119.869 -0.865190 8
  12  O1    os    M    9   8   5     1.703   104.602    -8.497 -0.468120 8
  13  C1    c3    M   12   9   8     1.416   117.733    68.314  0.158600 6
  14  H1    h1    E   13  12   9     1.108   110.793   -42.414  0.097080 1
  15  H2    h1    E   13  12   9     1.107   109.749    75.972  0.097080 1
  16  C2    c2    M   13  12   9     1.506   112.746  -162.419 -0.303240 6
  17  H3    ha    E   16  13  12     1.097   112.891  -157.710  0.087900 1
  18  C3    c2    M   16  13  12     1.347   130.091    29.798  0.124340 6
  19  C4    c3    3   18  16  13     1.517   123.815    -3.329 -0.835430 6
  20  H4    hc    E   19  18  16     1.103   110.653  -177.494  0.268710 1
  21  H5    hc    E   19  18  16     1.093   110.978   -55.946  0.268710 1
  22  H6    hc    E   19  18  16     1.104   111.342    60.720  0.268710 1
  23  C5    c3    M   18  16  13     1.518   119.219   171.116  0.035330 6
  24  H7    hc    E   23  18  16     1.107   110.434   131.394 -0.094760 1
  25  H8    hc    E   23  18  16     1.101   109.541    14.114 -0.094760 1
  26  C6    c3    M   23  18  16     1.563   114.428  -107.358  0.655040 6
  27  H9    hc    E   26  23  18     1.107   107.369  -178.136 -0.193200 1
  28  H10   hc    E   26  23  18     1.096   108.762   -63.269 -0.193200 1
  29  C7    c2    M   26  23  18     1.506   112.374    59.349 -0.792160 6
  30  H11   ha    E   29  26  23     1.095   114.503    59.298  0.202850 1
  31  C8    c2    M   29  26  23     1.352   127.382  -108.661  0.500260 6
  32  C9    c3    3   31  29  26     1.504   125.090    -7.524 -0.933960 6
  33  H12   hc    E   32  31  29     1.101   110.012  -138.154  0.272300 1
  34  H13   hc    E   32  31  29     1.097   113.244   -16.895  0.272300 1
  35  H14   hc    E   32  31  29     1.121   109.932   104.671  0.272300 1
  36  C10   c3    M   31  29  26     1.507   120.902   164.420 -0.933960 6
  37  H15   hc    E   36  31  29     1.102   111.644   132.401  0.272300 1
  38  H16   hc    E   36  31  29     1.100   111.905    10.156  0.272300 1
  39  H17   hc    E   36  31  29     1.113   107.363  -109.681  0.272300 1
```

```
LOOP

IMPROPER
   C3   C1   C2   H3
   C2   C4   C3   C5
   C8   C6   C7   H11
   C7   C10  C8   C9

DONE
```

(6) Parameters of OPP^{4-} (opp.prep)

```
    0    0    2

This is a remark line
molecule.res
OPP   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  O1    o     M    3   2   1     1.540   111.208  -180.000 -1.055690 8
   5  P1    p5    M    4   3   2     1.560   103.342    17.206  1.512520 15
   6  O2    o     E    5   4   3     1.560   112.594   -70.994 -1.055690 8
   7  O3    o     E    5   4   3     1.567   111.824    55.999 -1.055690 8
   8  O4    os    M    5   4   3     1.721   108.081   169.125 -0.690900 8
   9  P2    p5    M    8   5   4     1.722   156.994    67.207  1.512520 15
  10  O6    o     E    9   8   5     1.565   104.822  -127.691 -1.055690 8
  11  O7    o     E    9   8   5     1.557   109.672    -6.913 -1.055690 8
  12  O5    o     M    9   8   5     1.564   105.544   114.419 -1.055690 8


LOOP

IMPROPER

DONE
```

(7) Parameters of IM2 (int.prep)

```
    0    0    2

This is a remark line
molecule.res
INT   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
```

3	DUMM	DU	M	2	1	0	1.523	111.21	.0	.00000	
4	C9	c3	M	3	2	1	1.540	111.208	-180.000	-0.371600	6
5	H12	hc	E	4	3	2	1.098	121.750	-100.299	0.111360	1
6	H13	hc	E	4	3	2	1.098	131.609	87.417	0.111370	1
7	H14	hc	E	4	3	2	1.093	54.545	169.713	0.111370	1
8	C8	c2	M	4	3	2	1.509	59.796	-3.826	0.381790	6
9	C10	c3	3	8	4	3	1.510	114.373	-172.791	-0.371600	6
10	H15	hc	E	9	8	4	1.094	112.122	-179.701	0.111370	1
11	H16	hc	E	9	8	4	1.098	110.850	59.201	0.111370	1
12	H17	hc	E	9	8	4	1.098	110.814	-58.791	0.111370	1
13	C7	c2	M	8	4	3	1.342	125.263	7.367	-0.590250	6
14	H11	ha	E	13	8	4	1.092	117.365	-179.672	0.195110	1
15	C6	c3	M	13	8	4	1.508	128.052	-0.680	0.020020	6
16	H9	hc	E	15	13	8	1.108	109.371	118.513	0.068830	1
17	H10	hc	E	15	13	8	1.093	110.875	4.165	0.068830	1
18	C5	c3	M	15	13	8	1.548	111.272	-118.201	-0.326450	6
19	H7	hc	E	18	15	13	1.095	110.108	-40.822	0.166080	1
20	H8	hc	E	18	15	13	1.096	107.145	73.427	0.166080	1
21	C3	c3	M	18	15	13	1.562	118.642	-166.895	0.200520	6
22	C2	c2	B	21	18	15	1.523	107.187	-51.658	-0.101380	6
23	C1	c2	B	22	21	18	1.341	127.241	-115.888	-0.239750	6
24	H1	ha	E	23	22	21	1.086	122.766	-2.479	0.157340	1
25	H2	ha	E	23	22	21	1.087	120.835	176.006	0.157340	1
26	H3	ha	E	22	21	18	1.091	114.859	60.286	0.086550	1
27	C4	c3	3	21	18	15	1.538	107.759	-173.916	-0.664370	6
28	H4	hc	E	27	21	18	1.094	110.350	-64.715	0.191740	1
29	H5	hc	E	27	21	18	1.093	110.340	54.109	0.191740	1
30	H6	hc	E	27	21	18	1.096	112.885	174.551	0.191740	1
31	C21	c3	M	21	18	15	1.633	112.148	65.992	-0.008600	6
32	H28	hc	E	31	21	18	1.101	105.904	-73.641	0.074260	1
33	C20	c2	M	31	21	18	1.511	115.101	45.429	0.216630	6
34	O4	oh	S	33	31	21	1.352	119.285	-67.277	-0.582540	8
35	H27	ho	E	34	33	31	0.969	111.751	-6.902	0.473570	1
36	C19	c2	M	33	31	21	1.352	122.569	113.421	-0.201610	6
37	H26	ha	E	36	33	31	1.083	119.438	-178.504	0.228640	1
38	C18	ca	M	36	33	31	1.432	121.533	6.119	-0.087070	6
39	C13	ca	M	38	36	33	1.387	120.786	179.558	0.303280	6
40	O2	oh	S	39	38	36	1.353	119.654	3.380	-0.597130	8
41	H21	ho	E	40	39	38	0.975	107.327	177.902	0.448890	1
42	C12	ca	M	39	38	36	1.416	120.469	-174.632	0.136990	6
43	O1	os	S	42	39	38	1.359	113.423	179.661	-0.295160	8
44	C11	c3	3	43	42	39	1.448	119.278	-121.216	0.010960	6
45	H18	h1	E	44	43	42	1.094	109.873	57.071	0.091230	1
46	H19	h1	E	44	43	42	1.091	111.633	-65.457	0.091230	1
47	H20	h1	E	44	43	42	1.090	105.138	175.579	0.091230	1
48	C14	ca	M	42	39	38	1.408	123.103	-4.059	-0.032500	6
49	C15	c3	3	48	42	39	1.508	122.607	-172.254	-0.360530	6
50	H22	hc	E	49	48	42	1.092	110.018	145.534	0.137380	1
51	H23	hc	E	49	48	42	1.093	110.712	25.765	0.137380	1
52	H24	hc	E	49	48	42	1.097	112.125	-95.665	0.137380	1
53	C16	ca	M	48	42	39	1.396	117.374	4.446	0.391710	6
54	O3	oh	S	53	48	42	1.339	116.107	179.293	-0.532140	8

```
  55  H25   ho    E   54  53  48     0.976   110.520   179.581  0.452670 1
  56  C17   ca    M   53  48  42     1.456   121.171    -2.343 -0.314490 6
  57  C22   c2    M   56  53  48     1.386   123.129  -177.822  0.503440 6
  58  O5    oh    M   57  56  53     1.339   118.610    -3.566 -0.509140 8
  59  H29   ho    E   58  57  56     0.992   111.723   158.378  0.347520 1


LOOP
  C22  C21
  C17  C18

IMPROPER
   C7   C9   C8  C10
   C8   C6   C7  H11
   C1   C3   C2   H3
   C2   H1   C1   H2
  C19  C21  C20   O4
  C20  C18  C19  H26
  C19  C13  C18  C17
  C18  C12  C13   O2
  C13  C14  C12   O1
  C15  C12  C14  C16
  C14  C17  C16   O3
  C22  C18  C17  C16
  C21  C17  C22   O5

DONE
```

(8) Parameter of geranyl group (GER.prep)

```
    0    0    2

This is a remark line
molecule.res
GER   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  C1    c2    M    3   2   1     1.540   111.208  -180.000 -0.017260 6
   5  H1    ha    E    4   3   2     1.085   140.205   -96.695  0.185360 1
   6  H2    ha    E    4   3   2     1.087    71.654   154.007  0.185360 1
   7  C2    c2    M    4   3   2     1.359    64.776    15.574 -0.331420 6
   8  H3    ha    E    7   4   3     1.086   119.017    47.153  0.178870 1
   9  C3    c2    M    7   4   3     1.429   124.247  -135.165  0.524370 6
  10  C4    c3    3    9   7   4     1.485   123.612     7.344 -0.469260 6
  11  H4    hc    E   10   9   7     1.098   110.031   137.905  0.174020 1
  12  H5    hc    E   10   9   7     1.101   108.163  -107.585  0.174020 1
  13  H6    hc    E   10   9   7     1.090   115.034    14.071  0.174020 1
  14  C5    c3    M    9   7   4     1.460   118.309  -175.068 -0.271310 6
  15  H7    hc    E   14   9   7     1.093   111.362   160.657  0.134860 1
```

```
  16  H8    hc    E   14   9   7     1.094   112.516    36.488  0.134860 1
  17  C6    c3    M   14   9   7     1.630   109.297   -81.319  0.136470 6
  18  H9    hc    E   17  14   9     1.096   101.443  -158.245  0.055720 1
  19  H10   hc    E   17  14   9     1.091   109.219   -45.341  0.055720 1
  20  C7    c2    M   17  14   9     1.487   114.503    83.382 -0.507440 6
  21  H11   ha    E   20  17  14     1.091   114.634    62.400  0.196880 1
  22  C8    c2    M   20  17  14     1.352   127.667  -121.611  0.357200 6
  23  C10   c3    3   22  20  17     1.506   120.202  -174.937 -0.426270 6
  24  H15   hc    E   23  22  20     1.097   110.460   119.054  0.130250 1
  25  H16   hc    E   23  22  20     1.094   112.416    -1.784  0.130250 1
  26  H17   hc    E   23  22  20     1.099   110.882  -123.485  0.130250 1
  27  C9    c3    M   22  20  17     1.506   124.863     2.880 -0.426270 6
  28  H12   hc    E   27  22  20     1.099   110.038  -112.889  0.130250 1
  29  H13   hc    E   27  22  20     1.093   114.266     8.101  0.130250 1
  30  H14   hc    E   27  22  20     1.098   110.362   130.383  0.130250 1


LOOP

IMPROPER
   C2   H1   C1   H2
   C1   C3   C2   H3
   C2   C5   C3   C4
   C8   C6   C7  H11
   C7  C10   C8   C9

DONE
STOP
```

7 Reference

(1) Pertwee, R. G. (2006) Cannabinoid pharmacology: the first 66 years, *British Journal of Pharmacology 147*, S163-S171.

(2) Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L. M., Schwaiger, S., Heiss, E. H., Rollinger, J. M., Schuster, D., Breuss, J. M., Bochkov, V., Mihovilovic, M. D., Kopp, B., Bauer, R., Dirsch, V. M., and Stuppner, H. (2015) Discovery and resupply of pharmacologically active plant-derived natural products: A review, *Biotechnology Advances 33*, 1582-1614.

(3) Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015) The re-emergence of natural products for drug discovery in the genomics era, *Nature Reviews Drug Discovery 14*, 111-129.

(4) Newman, D. J., and Cragg, G. M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014, *Journal of Natural Products 79*, 629-661.

(5) Waltenberger, B., Mocan, A., Smejkal, K., Heiss, E. H., and Atanasov, A. G. (2016) Natural Products to Counteract the Epidemic of Cardiovascular and Metabolic Disorders, *Molecules 21*.

(6) Feher, M., and Schmidt, J. M. (2003) Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry, *Journal of Chemical Information and Computer Sciences 43*, 218-227.

(7) Samy, R. P., and Gopalakrishnakone, P. (2010) Therapeutic Potential of Plants as Anti-microbials for Drug Discovery, *Evidence-Based Complementary and Alternative Medicine 7*, 283-294.

(8) Newman, D. J., Cragg, G. M., and Snader, K. M. (2003) Natural products as sources of new drugs over the period 1981-2002, *Journal of Natural Products 66*, 1022-1037.

(9) Mishra, B. B., and Tiwari, V. K. (2011) Natural products: An evolving role in future drug discovery, *European Journal of Medicinal Chemistry 46*, 4769-4807.

(10) Fuloria, N. K., Raheja, R. K., Shah, K. H., Oza, M. J., Kulkarni, Y. A., Subramaniyan, V., Sekar, M., and Fuloria, S. (2022) Biological activities of meroterpenoids isolated from different sources, *Frontiers in Pharmacology 13*.

(11) Cornforth, J. W. (1968) Terpenoid biosynthesis, *Chemistry in Britain 4*, 102-106.

(12) Moss, G. P., Smith, P. A. S., and Tavernier, D. (1995) Glossary of Class Names of Organic-Compounds and Reactive Intermediates Based on Structure, *Pure and Applied Chemistry 67*, 1307-1375.

(13) Geris, R., and Simpson, T. J. (2009) Meroterpenoids produced by fungi, *Natural Product Reports 26*, 1063-1094.

(14) Shin-ya, K., Imai, S., Furihata, K., Hayakawa, Y., Kato, Y., Vanduyne, G. D., Clardy, J., and Seto, H. (1990) Isolation and structural elucidation of an antioxidative agent, naphterpin, *J Antibiot (Tokyo) 43*, 444-447.

(15) Ishibashi, M., Funayama, S., Anraku, Y., Komiyama, K., and Omura, S. (1991) Novel antibiotics, furaquinocins C, D, E, F, G and H., *Journal of Antibiotics 44*, 390-395.

(16) Shiomi, K., Nakamura, H., Iinuma, H., Naganawa, H., Isshiki, K., Takeuchi, T., Umezawa, H., and Iitaka, Y. (1986) Structures of new antibiotics napyradiomycins, *J Antibiot (Tokyo) 39*, 494-501.

(17) Pathirana, C., Jensen, P. R., and Fenical, W. (1992) Marinone and Debromomarinone - Antibiotic Sesquiterpenoid Naphthoquinones of a New Structure Class from a Marine Bacterium, *Tetrahedron Letters 33*, 7663-7666.

(18) Song, Y., Huang, H., Chen, Y., Ding, J., Zhang, Y., Sun, A., Zhang, W., and Ju, J. (2013) Cytotoxic and antibacterial marfuraquinocins from the deep South China Sea-derived Streptomyces niveus SCSIO 3406, *Journal of Natural Products 76*, 2263-2268.

(19) Matsuda, Y., and Abe, I. (2016) Biosynthesis of fungal meroterpenoids, *Natural Product Reports 33*, 26-53.

(20) Kumano, T., Tomita, T., Nishiyama, M., and Kuzuyama, T. (2010) Functional characterization of the promiscuous prenyltransferase responsible for furaquinocin biosynthesis: identification of a physiological polyketide substrate and its prenylated reaction products, *Journal of Biological Chemistry 285*, 39663-39671.

(21) Seo, Y., Park, K. E., Kim, Y. A., Lee, H. J., Yoo, J. S., Ahn, J. W., and Lee, B. J. (2006) Isolation of tetraprenyltoluquinols from the brown alga Sargassum thunbergii, *Chemical & Pharmaceutical Bulletin 54*, 1730-1733.

(22) Joung, E. J., Gwon, W. G., Shin, T., Jung, B. M., Choi, J., and Kim, H. R. (2017) Anti-inflammatory action of the ethanolic extract from Sargassum serratifolium on lipopolysaccharide-stimulated mouse peritoneal macrophages and identification of active components, *Journal of Applied Phycology 29*, 563-573.

(23) Joung, E. J., Lee, B., Gwon, W. G., Shin, T., Jung, B. M., Yoon, N. Y., Choi, J. S., Oh, C. W., and Kim, H. R. (2015) Sargaquinoic acid attenuates inflammatory responses by regulating NF-kappa B and Nrf2 pathways in lipopolysaccharide-stimulated RAW 264.7 cells, *International Immunopharmacology 29*, 693-700.

(24) Oh, S. J., Joung, E. J., Kwon, M. S., Lee, B., Utsuki, T., Oh, C. W., and Kim, H. R. (2016) Anti-Inflammatory Effect of Ethanolic Extract of Sargassum serratifolium in Lipopolysaccharide-Stimulated BV2 Microglial Cells, *Journal of Medicinal Food 19*, 1023-1031.

(25) Kim, S. N., Lee, W., Bae, G. U., and Kim, Y. K. (2012) Anti-diabetic and hypolipidemic effects of Sargassum yezoense in db/db mice, *Biochemical and Biophysical Research Communications 424*, 675-680.

(26) Tsang, C. K., and Kamei, Y. (2004) Sargaquinoic acid supports the survival of neuronal PC 12D cells in a nerve growth factor-independent manner, *European Journal of Pharmacology 488*, 11-18.

(27) El-Demerdash, A., Kumla, D., and Kijjoa, A. (2020) Chemical Diversity and Biological Activities of

Meroterpenoids from Marine Derived-Fungi: A Comprehensive Update, *Marine Drugs 18*.

(28) Zhao, M., Tang, Y. Q., Xie, J. W., Zhao, Z. X., and Cui, H. (2021) Meroterpenoids produced by fungi: Occurrence, structural diversity, biological activities, and their molecular targets, *European Journal of Medicinal Chemistry 209*.

(29) Nazir, M., Saleem, M., Tousif, M. I., Anwar, M. A., Surup, F., Ali, I., Wang, D. J., Mamadalieva, N. Z., Alshammari, E., Ashour, M. L., Ashour, A. M., Ahmed, I., Elizbit, Green, I. R., and Hussain, H. (2021) Meroterpenoids: A Comprehensive Update Insight on Structural Diversity and Biology, *Biomolecules 11*.

(30) Kuzuyama, T., Noel, J. P., and Richard, S. B. (2005) Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products, *Nature 435*, 983-987.

(31) Haagen, Y., Gluck, K., Fay, K., Kammerer, B., Gust, B., and Heide, L. (2006) A gene cluster for prenylated naphthoquinone and prenylated phenazine biosynthesis in Streptomyces cinnamonensis DSM 1042, *ChemBioChem 7*, 2016-2027.

(32) Winter, J. M., Moffitt, M. C., Zazopoulos, E., McAlpine, J. B., Dorrestein, P. C., and Moore, B. S. (2007) Molecular basis for chloronium-mediated meroterpene cyclization: cloning, sequencing, and heterologous expression of the napyradiomycin biosynthetic gene cluster, *Journal of Biological Chemistry 282*, 16362-16368.

(33) Kawasaki, T., Hayashi, Y., Kuzuyama, T., Furihata, K., Itoh, N., Seto, H., and Dairi, T. (2006) Biosynthesis of a natural polyketide-isoprenoid hybrid compound, furaquinocin A: identification and heterologous expression of the gene cluster, *Journal of Bacteriology 188*, 1236-1244.

(34) Noguchi, T., Zhao, F., Moriwaki, Y., Yamamoto, H., Nagata, R., Tomita, T., Terada, T., Shimizu, K., Nishiyama, M., and Kuzuyama, T. (2023) Reductive biosynthetic strategy for natural products via transient diazotization, *submitted for publication*.

(35) Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., and Horinouchi, S. (1999) A new pathway for polyketide synthesis in microorganisms, *Nature 400*, 897-899.

(36) Abdelraheem, E., Thair, B., Varela, R. F., Jockmann, E., Popadic, D., Hailes, H. C., Ward, J. M., Iribarren, A. M., Lewkowicz, E. S., Andexer, J. N., Hagedoorn, P. L., and Hanefeld, U. (2022) Methyltransferases: Functions and Applications, *ChemBioChem 23*, e202200212.

(37) Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence, *Trends Biochem Sci 28*, 329-335.

(38) Cheng, X. D., and Roberts, R. J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping, *Nucleic Acids Research 29*, 3784-3795.

(39) Liscombe, D. K., Louie, G. V., and Noel, J. P. (2012) Architectures, mechanisms and molecular evolution of natural product methyltransferases, *Natural Product Reports 29*, 1238-1250.

(40) Vidgren, J., Svensson, L. A., and Liljas, A. (1994) Crystal structure of catechol O-methyltransferase, *Nature 368*, 354-358.

(41) Dixon, M. M., Huang, S., Matthews, R. G., and Ludwig, M. (1996) The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12, *Structure 4*, 1263-1275.

(42) Raux, E., Schubert, H. L., Woodcock, S. C., Wilson, K. S., and Warren, M. J. (1998) Cobalamin (vitamin B12) biosynthesis--cloning, expression and crystallisation of the Bacillus megaterium S-adenosyl-L-methionine-dependent cobalt-precorrin-4 transmethylase CbiF, *European Journal of Biochemistry 254*, 341-346.

(43) Schubert, H. L., Wilson, K. S., Raux, E., Woodcock, S. C., and Warren, M. J. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase, *Nat Struct Biol 5*, 585-592.

(44) Krishnamohan, A., and Jackman, J. E. (2019) A Family Divided: Distinct Structural and Mechanistic Features of the SpoU-TrmD (SPOUT) Methyltransferase Superfamily, *Biochemistry 58*, 336-345.

(45) Liu, R. J., Zhou, M., Fang, Z. P., Wang, M., Zhou, X. L., and Wang, E. D. (2013) The tRNA recognition mechanism of the minimalist SPOUT methyltransferase, TrmL, *Nucleic Acids Research 41*, 7828-7842.

(46) Lim, K., Zhang, H., Tempczyk, A., Krajewski, W., Bonander, N., Toedt, J., Howard, A., Eisenstein, E., and Herzberg, O. (2003) Structure of the YibK methyltransferase from Haemophilus influenzae (HI0766): a cofactor bound at a site formed by a knot, *Proteins 51*, 56-67.

(47) Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., and Gamblin, S. J. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9, *Nature 421*, 652-656.

(48) Zhang, F., Chen, N., and Wu, R. (2016) Molecular Dynamics Simulations Elucidate Conformational Dynamics Responsible for the Cyclization Reaction in TEAS, *J Chem Inf Model 56*, 877-885.

(49) Kozbial, P. Z., and Mushegian, A. R. (2005) Natural history of S-adenosylmethionine-binding proteins, *BMC Structural Biology 5*, 19.

(50) Mahmoud, A. M., and Ali, M. M. (2019) Methyl Donor Micronutrients that Modify DNA Methylation and Cancer Outcome, *Nutrients 11*.

(51) Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., and McCann, P. P. (1996) S-adenosylmethionine and methylation, *FASEB Journal 10*, 471-480.

(52) Hegazi, M. F., Borchardt, R. T., and Schowen, R. L. (1976) SN2-LIKE TRANSITION-STATE FOR METHYL TRANSFER CATALYZED BY CATECHOL-O-METHYLTRANSFERASE, *Journal of the American Chemical Society 98*, 3048-3049.

(53) Sledziewskagojska, E. (1995) INACTIVATION OF O-6-METHYLGUANINE-DNA METHYLTRANSFERASE IN-VIVO BY SN2 ALKYLATING-AGENTS, *Mutation Research-DNA Repair 336*, 61-67.

(54) Poulin, M. B., Schneck, J. L., Matico, R. E., McDevitt, P. J., Huddleston, M. J., Hou, W. F., Johnson,

N. W., Thrall, S. H., Meek, T. D., and Schramm, V. L. (2016) Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36, *Proceedings of the National Academy of Sciences, USA 113*, 1197-1201.

(55) Linscott, J. A., Kapilashrami, K., Wang, Z., Senevirathne, C., Bothwell, I. R., Blum, G., and Luo, M. K. (2016) Kinetic isotope effects reveal early transition state of protein lysine methyltransferase SET8, *Proceedings of the National Academy of Sciences, USA 113*, E8369-E8378.

(56) Constable, D. J. C., Dunn, P. J., Hayler, J. D., Humphrey, G. R., Leazer, J. L., Linderman, R. J., Lorenz, K., Manley, J., Pearlman, B. A., Wells, A., Zaks, A., and Zhang, T. Y. (2007) Key green chemistry research areas - a perspective from pharmaceutical manufacturers, *Green Chemistry 9*, 411-420.

(57) Mao, N. D., Ye, Y., Zhuo, X. T., Ye, X. Y., and Xie, T. (2021) Recent advances of the site-specific direct methylation on aromatic rings, *Tetrahedron 96*.

(58) Pavkov-Keller, T., Steiner, K., Faber, M., Tengg, M., Schwab, H., Gruber-Khadjawi, M., and Gruber, K. (2017) Crystal Structure and Catalytic Mechanism of CouO, a Versatile C-Methyltransferase from Streptomyces rishiriensis, *PLoS One 12*.

(59) Tsutsumi, H., Moriwaki, Y., Terada, T., Shimizu, K., Shin-ya, K., Katsuyama, Y., and Ohnishi, Y. (2022) Structural and Molecular Basis of the Catalytic Mechanism of Geranyl Pyrophosphate C6-Methyltransferase: Creation of an Unprecedented Farnesyl Pyrophosphate C6-Methyltransferase, *Angewandte Chemie International Edition 61*.

(60) Rohmer, M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, *Natural Products Reports 16*, 565-574.

(61) Bloch, K. (1992) Sterol molecule: structure, biosynthesis, and function, *Steroids 57*, 378-383.

(62) Chang, H. Y., Cheng, T. H., and Wang, A. H. (2021) Structure, catalysis, and inhibition mechanism of prenyltransferase, *IUBMB Life 73*, 40-63.

(63) Pandian, S., Saengchjan, S., and Raman, T. S. (1981) An alternative pathway for the biosynthesis of isoprenoid compounds in bacteria, *Biochem J 196*, 675-681.

(64) Christianson, D. W. (2017) Structural and Chemical Biology of Terpenoid Cyclases, *Chemical Reviews 117*, 11570-11648.

(65) Tello, M., Kuzuyama, T., Heide, L., Noel, J. P., and Richard, S. B. (2008) The ABBA family of aromatic prenyltransferases: broadening natural product diversity, *Cellular and Molecular Life Sciences 65*, 1459-1463.

(66) Metzger, U., Keller, S., Stevenson, C. E., Heide, L., and Lawson, D. M. (2010) Structure and mechanism of the magnesium-independent aromatic prenyltransferase CloQ from the clorobiocin biosynthetic pathway, *Journal of Molecular Biology 404*, 611-626.

(67) Tudzynski, P., Holter, K., Correia, T., Arntz, C., Grammel, N., and Keller, U. (1999) Evidence for an ergot alkaloid gene cluster in Claviceps purpurea, *Molecular and General Genetics 261*, 133-141.

(68) Cheng, W., and Li, W. K. (2014) Structural Insights into Ubiquinone Biosynthesis in Membranes, *Science 343*, 878-881.

(69) Paldino, E., Cenciarelli, C., Giampaolo, A., Milazzo, L., Pescatori, M., Hassan, H. J., and Casalbore, P. (2014) Induction of Dopaminergic Neurons From Human Wharton's Jelly Mesenchymal Stem Cell by Forskolin, *Journal of Cellular Physiology 229*, 232-244.

(70) Tsutsumi, H., Moriwaki, Y., Terada, T., Shimizu, K., Shin-ya, K., Katsuyama, Y., and Ohnishi, Y. (2022) Structural and Molecular Basis of the Catalytic Mechanism of Geranyl Pyrophosphate C6-Methyltransferase: Creation of an Unprecedented Farnesyl Pyrophosphate C6-Methyltransferase, *Angewandte Chemie-International Edition 61*.

(71) Moriwaki, Y., Yato, M., Terada, T., Saito, S., Nukui, N., Iwasaki, T., Nishi, T., Kawaguchi, Y., Okamoto, K., Arakawa, T., Yamada, C., Fushinobu, S., and Shimizu, K. (2019) Understanding the Molecular Mechanism Underlying the High Catalytic Activity of p-Hydroxybenzoate Hydroxylase Mutants for Producing Gallic Acid, *Biochemistry 58*, 4543-4558.

(72) Arakawa, T., Sato, Y., Yamada, M., Takabe, J., Moriwaki, Y., Masamura, N., Kato, M., Aoyagi, M., Kamoi, T., Terada, T., Shimizu, K., Tsuge, N., Imai, S., and Fushinobu, S. (2020) Dissecting the Stereocontrolled Conversion of Short-Lived Sulfenic Acid by Lachrymatory Factor Synthase, *Acs Catalysis 10*, 9-19.

(73) Lim, K. J. H., Hartono, Y. D., Xue, B., Go, M. K., Fan, H., and Yew, W. S. (2022) Structure-Guided Engineering of Prenyltransferase NphB for High-Yield and Regioselective Cannabinoid Production, *Acs Catalysis 12*, 4628-4639.

(74) Garcia-Borras, M., Romero-Rivera, A., and Osuna, S. (2017) Computational tools for the evaluation of laboratory-engineered biocatalysts, *Abstracts of Papers of the American Chemical Society 253*.

(75) Sato, H., and Nakano, M. (2023) Concertedness and Activation Energy Control by Distal Methyl Group during Ring Contraction/Expansion in Scalarane-Type Sesterterpenoid Biosynthesis, *Chemistry 29*, e202203076.

(76) Pulido, A., Oliver-Tomas, B., Renz, M., Boronat, M., and Corma, A. (2013) Ketonic decarboxylation reaction mechanism: a combined experimental and DFT study, *ChemSusChem 6*, 141-151.

(77) Fomina, L., Vazquez, B., Tkatchouk, E., and Fomine, S. (2002) The Glaser reaction mechanism. A DFT study, *Tetrahedron 58*, 6741-6747.

(78) Ohashi, M., Liu, F., Hai, Y., Chen, M., Tang, M. C., Yang, Z., Sato, M., Watanabe, K., Houk, K. N., and Tang, Y. (2017) SAM-dependent enzyme-catalysed pericyclic reactions in natural product biosynthesis, *Nature 549*, 502-506.

(79) Levandowski, B. J., Zou, L., and Houk, K. N. (2018) Hyperconjugative Aromaticity and Antiaromaticity Control the Reactivities and pi-Facial Stereoselectivities of 5-Substituted Cyclopentadiene Diels-Alder Cycloadditions, *Journal of Organic Chemistry 83*, 14658-14666.

(80) Warshel, A., and Levitt, M. (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic

and steric stabilization of the carbonium ion in the reaction of lysozyme, *Journal of Molecular Biology 103*, 227-249.

(81) van der Kamp, M. W., and Mulholland, A. J. (2013) Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology, *Biochemistry 52*, 2708-2728.

(82) Mackerell, A. D., Jr. (2004) Empirical force fields for biological macromolecules: overview and issues, *Journal of Computational Chemistry 25*, 1584-1604.

(83) Dong, L., and Liu, Y. (2023) Catalytic Mechanism of Pyridoxal 5'-Phosphate-Dependent Aminodeoxychorismate Lyase: A Computational QM/MM Study, *J Chem Inf Model 63*, 1313-1322.

(84) Viegas, M. F., Neves, R. P. P., Ramos, M. J., and Fernandes, P. A. (2022) QM/MM Study of the Reaction Mechanism of Thermophilic Glucuronoyl Esterase for Biomass Treatment, *Chemphyschem 23*, e202200269.

(85) Wong, C. P., Awakawa, T., Nakashima, Y., Mori, T., Zhu, Q., Liu, X., and Abe, I. (2018) Two Distinct Substrate Binding Modes for the Normal and Reverse Prenylation of Hapalindoles by the Prenyltransferase AmbP3, *Angewandte Chemie, International Edition in English 57*, 560-563.

(86) Dill, K. A., and MacCallum, J. L. (2012) The protein-folding problem, 50 years on, *Science 338*, 1042-1046.

(87) Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2021) Critical assessment of methods of protein structure prediction (CASP)-Round XIV, *Proteins 89*, 1607-1617.

(88) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature 596*, 583-589.

(89) Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zidek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, *Nucleic Acids Research 50*, D439-D444.

(90) Hu, L., Salmen, W., Sankaran, B., Lasanajak, Y., Smith, D. F., Crawford, S. E., Estes, M. K., and Prasad, B. V. V. (2022) Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography, *Commun Biol 5*, 419.

(91) Hutin, S., Ling, W. L., Tarbouriech, N., Schoehn, G., Grimm, C., Fischer, U., and Burmeister, W. P. (2022) The Vaccinia Virus DNA Helicase Structure from Combined Single-Particle Cryo-Electron

Microscopy and AlphaFold2 Prediction, *Viruses 14*.

(92) Zhang, Y., Vass, M., Shi, D., Abualrous, E., Chambers, J. M., Chopra, N., Higgs, C., Kasavajhala, K., Li, H., Nandekar, P., Sato, H., Miller, E. B., Repasky, M. P., and Jerome, S. V. (2023) Benchmarking Refined and Unrefined AlphaFold2 Structures for Hit Discovery, *J Chem Inf Model 63*, 1656-1667.

(93) Groenhof, G. (2013) Introduction to QM/MM simulations, *Methods Mol Biol 924*, 43-66.

(94) Vreven, T., Byun, K. S., Komaromi, I., Dapprich, S., Montgomery, J. A., Morokuma, K., and Frisch, M. J. (2006) Combining quantum mechanical methods with molecular mechanics methods in ONIOM, *Journal of Chemical Theory and Computation 232*, 408-408.

(95) Yang, Z., Zeng, X., Zhao, Y., and Chen, R. (2023) AlphaFold2 and its applications in the fields of biology and medicine, *Signal Transduct Target Ther 8*, 115.

(96) Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution, *Journal of Computer-Aided Molecular Design 24*, 591-604.

(97) Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007) Epik: a software program for pK (a) prediction and protonation state generation for drug-like molecules, *Journal of Computer-Aided Molecular Design 21*, 681-691.

(98) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams, Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery Jr, J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., and Fox, D. J. (2016) Gaussian 16 Rev. C.01.

(99) Wang, J. M., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations, *Journal of Molecular Graphics and Modelling 25*, 247-260.

(100) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *Journal of Chemical Theory and Computation 11*, 3696-3713.

(101) Mobley, D. L., Bayly, C. I., Cooper, M. D., Shirts, M. R., and Dill, K. A. (2015) Correction to Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge

Atomistic Simulations (vol 5, pg 350, 2009), *Journal of Chemical Theory and Computation 11*, 1347-1347.

(102) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.

(103) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *Journal of Chemical Physics 79*, 926-935.

(104) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems *Journal of Chemical Physics 98*, 10089-10092.

(105) Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *Journal of Chemical Physics 23*, 327-341.

(106) Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics 81*, 3684-3690.

(107) Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Goetz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K. M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., Pan, F., Pantano, S., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Skrynnikov, N. R., Smith, J., Swails, J., Walker, R. C., Wang, J., Wilson, L., Wolf, R. M., Wu, X., Xiong, Y., Xue, Y., York, D. M., and Kollman, P. A. (2020) AMBER 2020.

(108) Maseras, F., and Morokuma, K. (1995) IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states, *Journal of Computational Chemistry 16*, 1170-1179.

(109) Dapprich, S., Komaromi, I., Byun, K. S., Morokuma, K., and Frisch, M. J. (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives, *Journal of Molecular Structure 461*, 1-21.

(110) Svensson, M., Humbel, S., Froese, R. D. J., Matsubara, T., Sieber, S., and Morokuma, K. (1996) ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels − Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition, *Journal of Physical Chemistry 100*, 19357-19363.

(111) Becke, A. D. (1993) DENSITY-FUNCTIONAL THERMOCHEMISTRY .3. THE ROLE OF EXACT EXCHANGE, *Journal of Chemical Physics 98*, 5648-5652.

(112) Lee, C. T., Yang, W. T., and Parr, R. G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Physical Review B: Condensed Matter 37*, 785-789.

(113) Tao, P., Fisher, J. F., Shi, Q. C., Vreven, T., Mobashery, S., and Schlegel, H. B. (2009) Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and Its Oxirane Analogue, *Biochemistry 48*, 9839-9847.

(114) Patel, P. D., and Masunov, A. E. (2009) Theoretical Study of Photochromic Compounds. 1. Bond Length Alternation and Absorption Spectra for the Open and Closed Forms of 29 Diarylethene Derivatives, *Journal of Physical Chemistry A 113*, 8409-8414.

(115) Ho, J. M., Klamt, A., and Coote, M. L. (2010) Comment on the Correct Use of Continuum Solvent Models, *Journal of Physical Chemistry A 114*, 13442-13444.

(116) Olsson, M. H., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions, *J Chem Theory Comput 7*, 525-537.

(117) Horberg, J., Saenz-Mendez, P., and Eriksson, L. A. (2018) QM/MM Studies of Dph5-A Promiscuous Methyltransferase in the Eukaryotic Biosynthetic Pathway of Diphthamide, *Journal of Chemical Information and Modeling 58*, 1406-1414.

(118) Chen, S. C., Huang, C. H., Lai, S. J., Liu, J. S., Fu, P. K., Tseng, S. T., Yang, C. S., Lai, M. C., Ko, T. P., and Chen, Y. (2015) Structure and mechanism of an antibiotics-synthesizing 3-hydroxykynurenine C-methyltransferase, *Scientific Reports 5*.

(119) Liptak, M. D., and Shields, G. C. (2001) Accurate pK(a) calculations for carboxylic acids using Complete Basis Set and Gaussian-n models combined with CPCM continuum solvation methods, *Journal of the American Chemical Society 123*, 7314-7319.

(120) Schonherr, H., and Cernak, T. (2013) Profound Methyl Effects in Drug Discovery and a Call for New C-H Methylation Reactions, *Angewandte Chemie International Edition 52*, 12256-12267.

(121) Wessjohann, L., Dippe, M., Tengg, M., and Gruber-Khadjawi, M. (2014) Methyltransferases in Biocatalysis, *Cascade Biocatalysis: Integrating Stereoselective and Environmentally Friendly Reactions*, 393-425.

(122) Yang, Y., Miao, Y., Wang, B., Cui, G., and Merz, K. M., Jr. (2012) Catalytic mechanism of aromatic prenylation by NphB, *Biochemistry 51*, 2606-2618.

(123) Mehmood, R., and Kulik, H. J. (2020) Both Configuration and QM Region Size Matter: Zinc Stability in QM/MM Models of DNA Methyltransferase, *J Chem Theory Comput 16*, 3121-3134.

(124) Yao, J., Chu, Y., An, R., and Guo, H. (2012) Understanding product specificity of protein lysine methyltransferases from QM/MM molecular dynamics and free energy simulations: the effects of mutation on SET7/9 beyond the Tyr/Phe switch, *J Chem Inf Model 52*, 449-456.

(125) Rod, T. H., and Ryde, U. (2005) Accurate QM/MM Free Energy Calculations of Enzyme Reactions: Methylation by Catechol O-Methyltransferase, *J Chem Theory Comput 1*, 1240-1251.

8 Acknowledgements

本論文の作成にあたり、多くの方々にご指導、ご鞭撻を賜りました。お世話になった皆様方に、この場で感謝の意を申し上げます。

本研究を遂行するにあたり、日頃から温かいご指導をいただき、また本論文の主査として貴重なご指摘、ご討論を賜りました本学大学院農学生命科学研究科　生物情報工学研究室 寺田透 准教授に心より感謝申し上げます。

元の指導教員　清水謙多郎　教授には、分子シミュレーションの分野で初学者である筆者を受け入れ、研究に関するあらゆることを教えていただきました。心より厚く御礼申し上げます。

本研究課題は、本学農学生命科学研究科 分子育種学研究室　葛山智久　教授、野口智弘博士（現　国立研究開発法人産業技術総合研究所）との共同研究により達成できました。深くお礼申し上げます。

研究に留まらず多岐にわたって有益なご指導を賜り、本論文の査読を行ってくださいました本研究室　森脇由隆　助教に心より感謝申し上げます。

また、日々の研究生活におきまして様々なご支援を賜りました生物情報研究室の学生の皆様に感謝いたします。

SPRING GX プログラム生として採用していただき、効率よく研究を遂行することができましたことに深くお礼を申し上げます。

最後に、あらゆる場合で筆者を支え、励ましてくれた家族と友人に心より感謝申し上げます。

2023 年 12 月

趙　凡