Computational studies of enzymatic reactions

Elucidating the Taxadiene Synthase mechanism and
the CalB-catalyzed hydrolysis of propranolol esters.

Inaugural-Dissertation

zur Erlangung des Doktorgrades der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

vorgelegt von (Jeaphianne) Jeannette Philomena Maria van Rijn

aus Bangkok - Thailand

Mülheim an der Ruhr, August 2020

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Heinrich-Heine-Universität Düsseldorf

Referent: Herr Prof. Dr. Frank Neese

Korreferentin: Frau Prof. Dr. Christel Marian

Tag der mündlichen Prüfung: 07.12.2020

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner Institution eingereicht. Ich habe keine erfolglosen Promotionsversuche unternommen.

Mülheim an der Ruhr, den 26. August 2020

(Jeannette P. M. van Rijn)

Abstract

Taxol (paclitaxel, an anticancer drug) and Propranolol (a beta-adrenergic blocking agent) are two important drugs for which insights into the reaction mechanism of the synthesis could open up the possibility to improve the production process through rational design of enzymes, mutations or better substrates.

For Taxol, the current commercial production processes heavily depend on the taxus plant and produce significant toxic waste streams, making them less environmentally sustainable and increasing the cost of taxol. The first committed step in the production of Taxol is the conversion of geranylgeranyl diphosphate (GGPP) to taxadiene (**T**), catalyzed by taxadiene synthase (TXS).
For this step we performed molecular dynamics (MD) simulations to study the dynamic behavior of noncovalent enzyme carbocation complexes. Taxadiene and the observed four side products originate from the deprotonation of carbocation intermediates. The MD simulations of the TXS carbocation complexes provide insights into potential deprotonation mechanisms of such carbocations, showing water bridges which may allow the formation of side products via multiple proton transfer reactions. Combined quantum mechanics/molecular mechanics (QM/MM) calculations were used to investigate energy profiles for the conversion of GGPP to **T**, as well as to minor products for different configurations of relevant TXS carbocation complexes. The QM/MM calculations suggest a reaction pathway for the conversion of GGPP to **T**, which slightly differs from previous proposals regarding the number of reaction steps and the conformation of the carbocations. The QM/MM results also indicate that formation of minor products via water-assisted deprotonation of the carbocations is highly exothermic, by about -7 to -23 kcal/mol. Curiously, however, the computed barriers and reaction energies indicate that the formation of some of the minor products is more facile than the formation of **T**. Thus, the present calculations provide detailed insights into possible reaction pathways and into the origin of the promiscuity of TXS, but they do not reproduce the product distribution observed experimentally.

Propranolol is commercially available as a racemic mixture, though the *S*-enantiomer is significantly more active as a drug than both the *R*-enantiomer and the racemic mixture, while the racemic mixture has been shown to cause severe side effects. A possible way to produce enantiomerically pure Propranolol is the use of lipase catalyzed hydrolysis reactions of ester compounds.
The CalB-catalyzed hydrolysis of propranolol esters was investigated for a range of acyl donors of various structures and different lengths to understand how the structure of the acyl donor affects the binding of the propranolol ester with CalB, the reactivity and the enantioselectivity.
Docking results suggests that acyl donors with branched alkyl chains are too sterically demanding to be reactive. Subsequent molecular simulations of the propranolol esters with linear chains suggest the reactivity of propranolol esters with shorter chains (O-acetyl-propranolol, **M0**) to be high compared to propranolol esters with longer ones (O-propanoyl-propranolol, **M1** and O-butanoyl-propranolol, **M2**). The MD simulations also suggest that the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol (**M0**) will prefer to generate *R*-propranolol.
In agreement with MD results, the QM/MM calculations of the hydrolysis reaction of **M0** suggests an enantiomeric preference for the *R*-product. The activation energy gap between the reaction of ***R***- and ***S*-M0** is 6.2 kcal/mol, which is larger than that of the acylation reaction, indicating a potentially higher enantioselectivity for the hydrolysis reaction.

iv

Zusammenfassung

Taxol (Paclitaxel, ein Krebsmedikament) und Propranolol (ein Beta-Blocker) sind zwei wichtige Medikamente deren industrielle Synthesen durch ein detailliertes Verständnis der einzelnen Reaktionsschritte möglicherweise verbessert werden kann, beispielsweise durch das rationale Design von Enzymen zur Katalyse der einzelnen Reaktionsschritte, deren Mutation oder durch die Verwendung geeigneterer Substrate.

Für Taxol hängen derzeit die kommerziellen Produktionsprozesse stark von der Gewinnung aus der Rinde der Pazifischen Eibe ab und erzeugen erhebliche Mengen an Umweltgiften. Diese Prozesse sind daher nicht umweltverträglich und erhöhen außerdem die Kosten von Taxol. Der erste festgelegte Schritt bei der Herstellung von Taxol ist die Umwandlung von Geranylgeranyldiphosphat (GGPP) in Taxadien (**T**), welcher durch die Taxadiensynthase (TXS) katalysiert wird.

Für diesen Reaktionsschirtt wurden Molekulardynamik-(MD)-Simulationen durchgeführt um das dynamische Verhalten des Komplexes aus Enzym und des involvierten Carbokations zu untersuchen. Taxadien und die vier beobachteten Nebenprodukte des ersten Reaktionsschritts stammen jeweils aus der Deprotonierung des simulierten Carbokationzwischenprodukts. Die MD-Simulationen der TXS-Carbokationkomplexe liefern hierbei Einblicke in die jeweilgen Deprotonierungsmechanismen des Carbokations. Die Ergebnisse der MD-Simulationen unterstützen eine frühere Hypothese nicht, welche den Taumelbewegungen des Carbokations eine Schlüsselrolle bei der Deprotonierung durch das Pyrophosphat zuschreibt. Stattdessen werden Wasserstoffbrückenbindungen identifiziert, welche die Bildung von Nebenprodukten durch mehrere Protonentransferreaktionen ermöglichen.

Anhand von kombinierten quantenmechanischen/molekülmechanischen (QM/MM) Rechnungen wurden die Energieprofile der Reaktionen für die Umwandlung von GGPP in **T** sowie in die Nebenprodukte für verschiedene Konfigurationen des TXS-Carbokationkomplexes untersucht. Die Ergebnisse der QM/MM-Rechnungen legen einen Reaktionsweg für die Umwandlung von GGPP in **T** nahe, der sich geringfügig von früheren Vorschlägen hinsichtlich der Anzahl der Reaktionsschritte und der Konformation des Carbokations unterscheidet. Die Ergebnisse der QM/MM-Rechnungen zeigen auch, dass die Bildung der Nebenprodukte durch wassergestützte Deprotonierung der Carbokationen mit einer Reaktionsenergie von -7 bis -23 kcal/mol stark exotherm ist.

Eigentümlicherweise deuten die berechneten Barrieren und Reaktionsenergien darauf hin, dass die Bildung einiger Nebenprodukte einfacher ist als die Bildung von **T**. Somit liefern die vorliegenden Rechnungen detaillierte Einblicke in mögliche Reaktionswege sowie in den Ursprung der Promiskuität von TXS, reproduzieren aber die experimentell beobachtete Produktverteilung nicht.

Propranolol ist im Handel als racemische Mischung der *S*- und *R*-Enantiomere erhältlich, obwohl das *S*-Enantiomer in pharmakologischer Hinsicht signifikant aktiver ist als das *R*-Enantiomer und das racemische Gemisch. Darüber hinaus verursacht die racemische Mischung schwerwiegende Nebenwirkungen. Ein möglicher Weg zur Herstellung von enantiomeren-reinem Propranolol beinhaltet die Verwendung von Lipase-katalysierten Hydrolysereaktionen von Esterverbindungen.

Die CalB-katalysierte Hydrolyse von Propranololester wurde für eine Reihe von Acyldonoren mit unterschiedlicher Struktur und Länge untersucht, um zu verstehen wie die chemische Struktur des Acyldonors die Bindung des Propranololesters an CalB, die Reaktivität und die Enantioselektivität beeinflusst.

Ergebnisse von Docking-Simulationen legen nahe, dass Acyldonoren mit verzweigten Alkylketten sterisch zu anspruchsvoll sind um reaktiv zu sein. Nachfolgende molekulare Simulationen der Propranololester mit linearen Ketten deuten darüber hinaus an, dass die Reaktivität von Propranololestern mit kürzeren Ketten (O-Acetyl-Propranolol, **M0**) im Vergleich zu Propranololestern mit längeren Ketten (O-Propanoyl-Propranolol, **M1** und O-Butanoyl-Propranolol, **M2**) hoch ist. Die MD-Simulationen sagen weiterhin voraus, dass die Hydrolysereaktion von racemischem (*R,S*)-O-Acetyl-Propranolol die Bildung von *R*-Propranolol begünstigt.

In übereinstimmung mit den Ergebnissen der MD-Simulationen, legen die QM/MM-Rechnungen der Hydrolysereaktion von **M0** eine Präferenz für das *R*-Enantiomer als Produkt nahe. Der Unterschied der Aktivierungsenergien zwischen der Reaktion des ***R*-** und ***S*-M0** beträgt 6,2 kcal/mol und ist damit größer als bei der Acylierungsreaktion. Letzteres weist auf eine möglicherweise erhöhte Enantioselektivität der Hydrolysereaktion hin.

Acknowledgements

Looking back I would have to say the past 6.5 years, with my PhD and everything that has happened outside of that, has been the most character-shaping period of my life. Naturally this means there are many people who I would like to thank for their contributions.

First I would like to thank Prof. Dr. Walter Thiel, for giving me the opportunity to work in his group. For being the most patient person I have ever met, for believing my projects would turn out okay and for helping to actually make that happen. On August 23, 2019 Prof. Thiel unexpectedly passed away, leaving us all in shock over the loss of this knowledgable and kind man.

In this unusual situation, Prof. Dr. Frank Neese agreed to be my supervisor for the final part of my PhD, for which I am very grateful. I would like to thank Prof. Dr. Christel Marian for helping me find a way to finish my PhD, for reviewing this thesis and for accepting to be my copromotor. Thanks to the members of the committee for reading and approving this thesis. I would like to thank the Max-Plack-Gesellschaft for funding my PhD studies and enabling me to attend summer schools and conferences; and the support staff of the MPI, especially Frau Scheifachen and Herr Lenk, for all their help.
I would also like to thank Prof. Dr. Thomas Brück for proposing the collaboration on the Taxadiene Synthase Catalysis project and Dr. Patrick Schrepfer for providing the necessary structures.

Academia is an environment where you meet many amazing people from all over the world, with whom you can build friendships easily. Unfortunately, academia is also an environment where you have to say goodbye to these friends.
I would like to express my gratitude to Dr. Andrès Escorcia, who joined our group at the perfect time and who, with his fresh perspective, his knowledge and his incredible eye for detail has really helped to successfully finish the projects. Together with AP Dr. Guijuan Cheng we have had a very productive time. I would also like to thank Andrès and Guijuan for enlivening my stays in Germany with fun and more serious conversations.
Another person who I am very grateful to is Dr. Matthias Heyden, who convinced me countless times to persevere, not just with my PhD, but also with learning Python. Thanks for making your sofa available, for our fun ski trip and for being a great friend. I also want to thank Matthias for correcting my crappy German translation of the abstract.
I would like to thank current and former colleagues in Germany. Thanks to the people from the group of Prof. List for inviting me to their lively parties and Fabio, thanks for showing me Rome. Thanks to Alec, Tatiana, Lina, Sergio, Tommaso, Angelo and all the others for making the European Summerschool in Quantum Chemistry an amazing experience with so much learning and so much fun. Sergio, thanks for later showing me Pisa.

This doctoral thesis is a publication-based thesis. The works on Taxadiene Synthase Catalysis as presented in this thesis have been published in international peer-reviewed scientific journals. A short summary of the published papers as well as a description of my contribution to each piece of work is given in Chapter 3.1 and 3.2. The two published papers are included as attachments to this thesis. Chapter 3.3 presents a study on CalB-catalyzed Hydrolysis which has not been published yet, but for which we intend to submit the manuscript in the near future.

Just keep swimming.

– Dory, Finding Nemo

List of publications

1   Molecular Dynamics Study of Taxadiene Synthase Catalysis

Andrés M. Escorcia, **Jeaphianne P. M. van Rijn**, Gui-Juan Cheng, Patrick Schrepfer, Thomas B. Bruck and Walter Thiel

Journal of Computational Chemistry, **2018**, 39, 1215–1225.

2   QM/MM Study of Taxadiene Synthase Catalysis

Jeaphianne P. M. van Rijn, Andrés M. Escorcia and Walter Thiel

Journal of Computational Chemistry, **2019**, 40, 1902–1910.

3   MD and QM/MM Studies on CalB-catalyzed Hydrolysis of Propanolol Esters: the Effects of Acyl Moiety.

Gui-Juan Cheng, Andrés M. Escorcia, **Jeaphianne P. M. van Rijn** and Walter Thiel

x, **2021 (in preparation)**, x, x–x.

Contents

1. Introduction

With the vast amount of drugs that has been discovered, many illnesses can nowadays be managed or cured. However, some important drugs are still not available in the desired quantity or quality. The utilization of biocatalysis is recognized as a potential solution.[1–6]

In biocatalysis, an enzyme is used to carry out (part of) the reaction, which has some great benefits: because enzymes are known to significantly increase the rate of a specific reaction, the reaction can take place at milder conditions, and because they are often very selective less raw materials are needed, making the entire process more sustainable and environmentally friendly.[4,7–9] With the trend of moving towards green chemistry and the improvement in the supply and modification of enzymes, the use of engineered enzymes in the production of pharmaceuticals is a logical step.[10]

Two important drugs for which the utilization of enzymes in the production process can be of great benefit are Taxol (paclitaxel), an anticancer drug[2,11,12] and Propranolol, a beta-adrenergic blocking agent used for the treatment of cardiovascular disorders.[13–15]

1.1. Taxol

1.1.1. The taxol production process

Researchers have been interested in the generic drug paclitaxel because of its great anticancer activity. Paclitaxel is commonly known by the registered trade name Taxol® (Bristol-Myers Squibb)[16] and it is approved for treatment of breast, lung and ovarian cancer[2,11,12,17] with new clinical applications anticipated.[2,18,19] The growing demand for the drug makes the taxol production process an interesting target for optimization.[2,10,20]

In 1962 taxol was initially extracted from the bark of the Pacific yew tree (*Taxus brevifolia*).[12] However, the taxol content of the bark is only 0.0001% to 0.08%, and the average yield of isolation is in the range of 0.014-0.017%.[11,21,22] In other words, about 7,000 kg of Taxus bark is needed to produce 1 kg of paclitaxel.[2,23] Moreover, during the extraction process, the slow growing and sparsely distributed Yew tree has to be destroyed, so researchers started looking

for a different way to produce taxol.[2]

In 1994 the labs of Holton et al. and Nicolaou et al.[24–26] reported two different pathways for the total synthesis of taxol. However, these and other pathways that have since been published,[27–32] have limited applicability because of their extreme complexity, toxic side products and low yields.[1, 2, 10]

Instead of using the bark, researchers shifted their attention to the renewable needles of the Yew trees. Though the needles generally contain less taxol than the bark, they were found to be a relatively good source (about 0.1%) of taxol intermediates baccatin III or 10-deacetylbaccatin III.[2, 11] From these precursors a semi-synthetic process is employed for production of taxol and its analog taxotere.[2, 33] Though using the needles makes the process less destructive, it still depends on the slow growing Taxus trees, in which the concentration of the intermediates fluctuates due to environmental factors.[2] Moreover, the necessary purification of these intermediates from the plant source is expensive.[2, 34]

Some steps of the paclitaxel biosynthetic pathway have been transferred into heterologous expression systems, but this was not possible for all steps in taxol production.[1, 10] The use of Taxus plant cell cultures is seen as a promising route, though this process has issues like unstable yields, low growth rates compared to bacterial cultures[35] and difficulties with extraction of the desired compound due to impurities.[36]

Currently, the commercial production of taxol is done using semi-synthesis and cell cultures.[10] Both routes still heavily depend on the Taxus plant and produce significant toxic waste streams, making them less environmentally sustainable and increasing the cost of paclitaxel.[1, 2, 10, 37]

In an effort to improve the production process, over the years, many steps in the biosynthetic taxol pathway from the universal acyclic diterpene precursor geranylgeranylpyrophosphate (GGPP, C20) have been identified.[1, 2, 38, 39] The first committed step in the biosynthetic pathway from GGPP to taxol is the conversion of GGPP into taxa-4(5),11(12)- diene (taxadiene).[40] This conversion is catalyzed by Taxadiene Synthase (TXS), an enzyme found in the *Taxus brevifolia* tree. Optimization of this step might be worthwhile, because taxadiene production in TXS is conceivably slow compared to the oxygenations steps further on in the taxol pathway, indicated by the low catalytic activity of TXS in plants.[10, 40–42] We will therefore focus on the cyclization of GGPP to taxadiene for the remainder of this work.
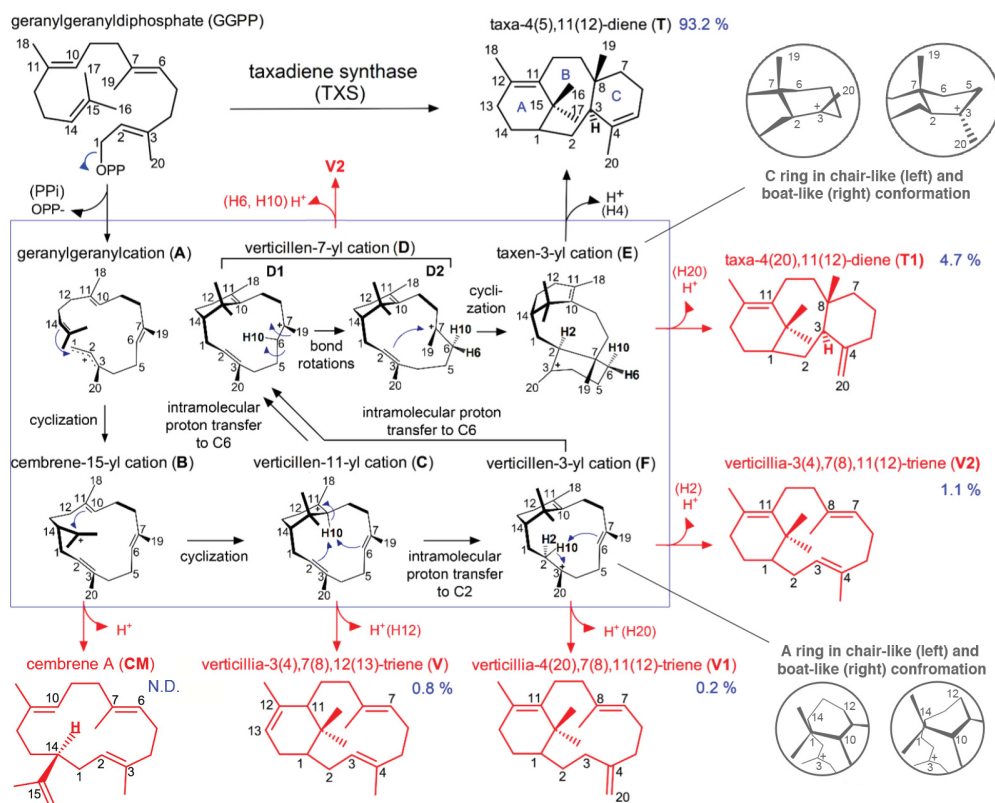
Figure 1.1.: In black: the reaction mechanism of the TXS-catalyzed cyclization of GGPP to taxadiene from QM model calculations (HT- QM mechanism).[38] In red: minor products of TXS catalysis. In blue: the WT product yields as reported by Schrepfer et al.[20] and the taxadiene ring systems labels (A-C). In gray: illustration of chair- and boat-like conformations of the A and C ring. For the carbocation intermediates the numbering of GGPP is used, while minor products are numbered according to taxadiene convention.

1.1.2. First committed step: GGPP to taxadiene

TXS initiates the catalysis using a trinuclear cluster of Mg^{2+} ions to bind GGPP and to trigger departure (ionization) of the pyrophosphate group (PPi).[43,44] This leads to formation of a highly reactive carbocation, which subsequently undergoes rearrangement and cyclization reactions to form the product (see figure 1.1).[43,44]

Though it has not been possible to experimentally trap GGPP-derived carbocation intermediates, labeling experiments suggest the existence of four cationic intermediates (**B**, **C**, **D**, **E**, figure 1.1).[45−48] Quantum mechanical (QM) studies complemented these experimental findings to indicate two more potential carbocation intermediates, **A** and **F**.[38,49]

Combined, these studies lead to the proposed reaction mechanism of taxadiene formation, here referred to as the HT-QM mechanism (HT being a label for its authors,[38] see figure 1.1), which includes the following steps: starting from GGPP, cation **A** is formed by pyrophosphate loss. Cyclization by C1-C14 σ-bond formation leads to cation **B** and a second ring closure between C15 and C10 yields cation **C**. Based on the experiments and calculations, it was proposed that the GGPP - **A** - **B** - **C** transformation might not be a fully concerted process, but that the lifetime for cation **B** will be very short.[38,46,47] From cation **C**, an intramolecular proton transfer from C10 to C7 leads to formation of **D**. In addition to this pathway, a two-step process (via formation of cation **F**) was proposed.[49] After a conformational change of **D**, a final cyclization leads to formation of cation **E**, which can be deprotonated at C4 to form taxadiene (**T**).

In addition to the main product **T**, the promiscuous TXS[20,46,48,50] produces several minor products like taxa-4(20),11(12)-diene (**T1**), verticillia-3(4),7(8),12(13)-triene (**V**), verticillia-4(20),7(8),11(12)-triene (**V1**) and verticillia-3(4),7(8), 11(12)-triene (**V2**) (see figure 1.1).[20] Though several studies report similar side products, the product distributions reported in literature vary greatly.[20,45,46,51] In this work we refer to the product yields as reported by Schrepfer et al. (see figure 1.1 in blue).[20]

Experiments carried out on the nonenzymatic chemical transformation of verticillene and its derivatives to taxadiene were unsuccessful or had a very low yield (0.004%).[52,53] This indicates that Taxadiene Synthase is critical for taxadiene formation.

1.1.3. Taxadiene Synthase

The full-length form of Taxadiene Synthase (TXS) contains 862 residues; including an N-terminal transit sequence of roughly 80 residues, which is cleaved on plastidial protein maturation.[54,55]

There are two classes of terpenoid cyclases, class I and class II, which use different substrate activation mechanisms and have different protein folds.[43,56–59] The catalytic TXS structure assembles into three α-helical domains, which harbor the folds of both class I and class II terpenoid cyclases.[43]

The C-terminal domain (S553-V862) has the class I fold. Typical for a class I terpenoid cyclase is the binding and activation of the substrate (GGPP) with a three-metal ion cluster (referred to as the Mg_A^{2+}, Mg_B^{2+} and Mg_C^{2+} ions). This metal cluster is ligated by conserved metal-binding motifs, which are found in TXS as D^{613}DMAD (at helix D) and N^{757}DTKTYQAE (at helix H).[43,56,60] The trinuclear metal cluster is needed to bind the ligand and to trigger ionization of the GGPP diphosphate group (PPi) to generate a carbocation, thereby initiating catalysis.

The N-terminal domain of TXS (M107-I135 and S349-Q552) and the "insertion" domain[61] (S136-Y348) adopt the fold of a class II terpenoid cyclase.[43] Whereas class I cyclases ionize the substrate, class II cyclases activate the substrate by protonating an isoprenoid double bond or oxirane moiety using the second aspartic acid in a DXDD motif.[43]

Taxadiene synthase does not contain a DXDD motif, but requires Mg^{2+} for optimal catalytic activity[41] and it does have the conserved metal-binding motifs. This and the fact that the active site (the exclusive binding site of the substrate) of TXS is located in the C-terminal domain indicates that TXS functions as a class I terpenoid cyclase.[43]

TXS in closed conformation

Upon binding of the substrate and the catalytic metal ions, the enzyme TXS undergoes a conformational change from an open to a closed conformation.[43,58,62-65] Closure of the active site protects the highly reactive carbocation intermediates from premature quenching by bulk solvent.[56,64] Other than that, the TXS enzyme has mainly been suggested to enforce the right conformation for the reaction,[58,62] though recent studies indicate electrostatic involvement.[20,62,63,66-73] To fully understand the enzymatic TXS reaction mechanism the protein-solvent environment needs to be taken into account.[63,74-77] However, in the available crystal structure of TXS (Protein Data Bank (PDB) ID 3P5R, which contains a fluorinated substrate analog (2-F-GGPP) and is missing the N-terminal), the active site is not fully closed.[43]

Considering the closed structure of other terpene synthases it is expected that in the closed conformation the N terminus random coil segment (hereafter referred to as the NTRC, D80-L108) caps the active site, while the A–C (G570-H579), J–K (F837-E846), and H−α1 (Q770-S773) loops and the C-terminal portion of helix H (K760-G769) flank the mouth of the active site (see figure 1.2).[20,43,54,72]

In the crystal structure, hydrogen bonds are found between the substrate and R754 and N757, and there are water-mediated hydrogen bonds with Y688, E691, Y835, S713, R768 and Q770.[43] In addition, the magnesium ions are bound to the conserved metal-binding motifs: the Mg_A^{2+} and Mg_C^{2+} ions are linked to D613 and D617, while N757, T761, and E765 coordinate to the Mg_B^{2+} ion (see figure 1.2).[43] In the closed conformation, the new orientation of the aforementioned loops would lead to formation of additional interactions between substrate and enzyme, as well as between the loops.

Unfortunately, crystal structures of closed TXS complexed with GGPP or carbocation analogues have not been reported. However, by applying homology modeling, two models of TXS in the closed conformation have been constructed up to now, which are referred to here as the SHM[20] and the FHM[72] (i.e. the Schrepfer and Freud Homology Models).
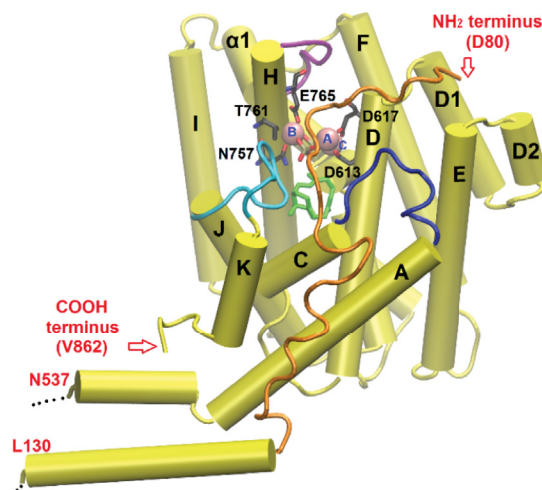
Figure 1.2.: Representative structure of the SHM of TXS in the closed active site conformation. Atomic coordinates were provided by Schrepfer et al.[20] The protein regions capping the active site are shown in different colors: loops A–C in blue, J–K in cyan, and H–α1 in purple. The N terminus random coil (NTRC) segment is shown in orange, while the C-terminal portion of helix H is shown in yellow like the rest of the protein. In licorice: residues acting as metal ligands (carbons in black), cation C (green), and PPi (phosphorus in tan). The catalytic Mg^{2+} ions are shown as spheres with their respective A-C labels in blue. Hydrogen atoms are omitted for clarity. The labels for helical segments are based on the farnesyl diphosphate synthase convention.[62, 78]

SHM   To obtain the SHM, closed bornyl diphosphate synthase (BPPS, PDB ID 1N20)[62] was used as a template for homology modeling followed by an energy minimization and a 10-ns MD simulation of the structure in water.[20, 43, 62] Models of the closed TXS complex, containing GGPP as well as all the carbocations in the HT-QM mechanism (hereafter denoted as $^{\textbf{TXS}}$**cation** complexes), were created using an integrated approach of docking and molecular mechanics (MM) optimizations.[20] Each of these $^{\textbf{TXS}}$**cation** complexes can be distinctly derived from its precursor in the reaction cascade (figure 1.1), indicating that the HT-QM mechanism is feasible in the TXS environment.[20] The product distribution for the wild type (WT) enzyme was also reported and rationalized.[20]

In the $^{\textbf{TXS}}$**GGPP** complex, an extended hydrogen bond network is found compared to the open conformation (see figure 1.2). In the SHM, PPi accepts water-mediated hydrogen bonds extending from R580 and N-terminal Y89 in the closed complex in a network similar to one found in BBPS.[20, 62] Based on the SHM, mutations were proposed and the predictive power of the SHM was confirmed by experimental site-directed mutagenesis data.[20] The studies on TXS catalysis in this work use the SHM of Taxadiene Synthase.

FHM For the construction of the FHM, sequence alignment, structure superposition, homology modeling, and loop modeling were employed using the closed form of bornyl diphosphate synthase from *Salvia officinalis* (BPPS, PDB ID 1N23[62]) as a template while also considering the structures of 5-epi-aristolochene synthase from *Nicotiana tabacum* (PDB code 3M02[65]) and limonene synthase from *Mentha spicata* (PDB code 2ONG[70]).[72]

The FHM was used to run hybrid QM/MM and free energy simulations.[72,73] These suggest a six step pathway for the biosynthesis of taxadiene from GGPP, where, in contrast to the HT-QM mechanism (figure 1.1), the direct pathway from **C** to **D1** is preferred over the indirect path via cation **F**.[72,73] The reaction was found to be strongly exothermic and downhill from cation **A** to taxadiene. A similar reaction profile has been reported for other terpene synthases.[63,76] A global free energy barrier of 24 kcal/mol was obtained, in good agreement with the experimentally determined k_{cat} of 0.0106 s^{-1} (ΔG$^{\ddagger} \approx$ 20.1 kcal/mol).[73] The reaction energetics for formation of side products was not addressed, although a qualitative explanation for the promiscuity of TXS was provided.[72,73]

In the FHM active-site architecture, residue W753 was identified to be critical for stabilizing the carbocations through π-cation interactions.[73] The TXS W753H mutant exclusively produces cembrene A **CM**,[20] which could be rationalized based on the FHM, further supporting the quality of this model.[73]

Differences SHM and FHM Despite being derived from a similar procedure based on the same template, the structures of the two active enzyme models, SHM and FHM, differ significantly, especially in the following three points.

1. The number of water molecules:
   While the setup with the FHM contains a single active-site water molecule, the setup with the SHM has 4 water molecules within 3 Å of the carbocation.

2. The orientation of the substrate in the binding pocket:
   In the FHM the positive center at $^{\textbf{TXS}}$**C**:C11 forms a π-cation interaction with Y841[73] (no distance reported, see figure 1.3-left), while in the SHM the C11$^{+}\cdots$Y841 distance is 5.8 Å (see figure 1.3-right), so the π-cation interaction is very weak. For more information see appendix B section S5 and figure S12.

3. The structure of the binding pocket:
   Both R578 (A-C loop) and D839 (J-K loop) are located over 12.5 Å away from PPi in the SHM, while the former residue interacts directly with the PPi moiety in the FHM and the latter makes a water-mediated hydrogen bond with PPi. Therefore it appears that
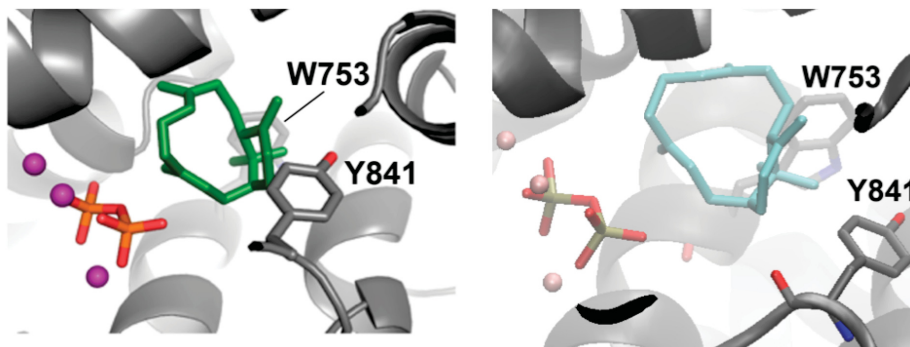
Figure 1.3.: The plot on the left shows cation **C** in the FHM as depicted in figure 2c of Ansbacher et al.;[73] residue labels have been added. The plot on the right depicts cation **C** in the SHM (setup **TXS**C:W1E1). In both plots water molecules are omitted for clarity.

the A-C (G570-H579) and J-K (F837-E846) loops are positioned differently leading to different active-site architectures. This can most likely be attributed to differences in the sequence alignment of the Arg-containing N-terminal region.

Elucidating the TXS mechanism   As indicated before, since the conversion of GGPP to taxadiene, which is the first committed step in the production of taxol, may be slow compared to the oxygenations steps further on in the taxol pathway, optimization of this step might be worthwhile in an attempt to improve the taxol production process. A more detailed understanding of TXS catalysis would be a first step towards eliminating the need for serendipity in such an optimization process.

This thesis presents two studies, which aim to gain a more detailed understanding of TXS catalysis based on the SHM and provide new insights into the molecular basis of TXS promiscuity.

Another drug for which the utilization of enzymes in the production process can be of great value is Propranolol.

1.2. Propranolol

(*R,S*)-1-iso-propylamino-3-(1-naphthoxy)-2-propanol (Propranolol) is a beta-adrenergic blocking agent, used for treatment of arterial hypertension and other cardiovascular disorders.[13–15] It is commercially available as racemic mixture, but the *S*-enantiomer is 60-100 times more active than the *R*-enantiomer and more potent than the racemic mixture.[15,79] Moreover,
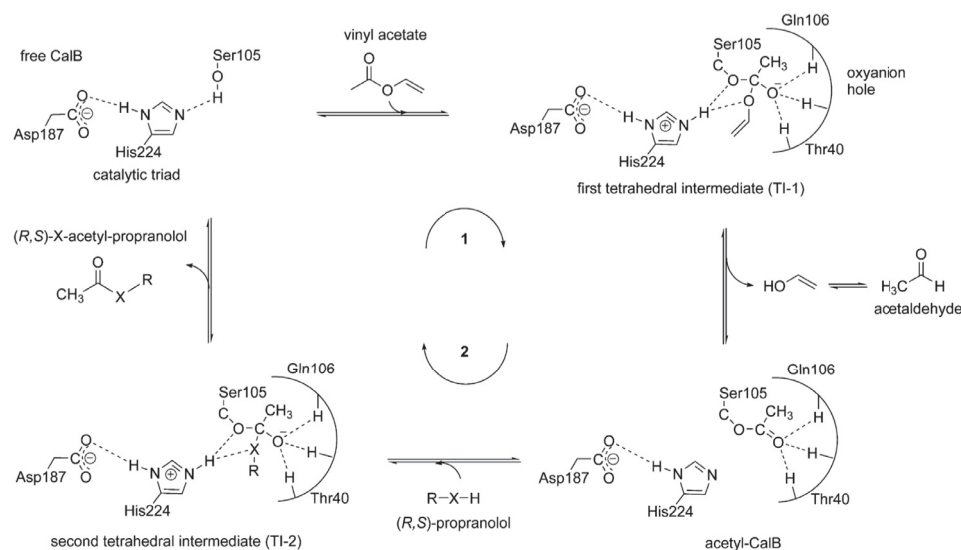
Figure 1.4.: Generally accepted reaction mechanism for lipase-catalyzed reactions. The case of the CalB-catalyzed acetylation of (*R,S*)-propranolol is shown as an example. The reaction consists of the acylation of CalB (1) followed by deacylation of CalB (2). Depending on the nucleophilic group of propranolol X can be N or O.

the racemic mixture has been shown to cause bronchoconstriction in humans,[80] while *in vivo* assays with rats indicate that racemic propranolol is more toxic than either of its pure enantiomers.[81, 82] Therefore, the development of methods to obtain enantiomerically pure *S*-propranolol is important and several strategies have been explored, including chemical, enzymatic and chemoenzymatic synthesis.[83–94]

1.2.1. Lipases

A well-established and versatile method to access enantiomerically pure compounds is the use of lipases.[3–6] Lipases are a class of enzymes that catalyze several reactions that involve carboxylic groups, like esterification, transesterification, aminolysis and hydrolysis.[95–99]

These enzymes can transform the enantiomers of a racemic mixture with different rates (kinetic resolution) and have been successfully used in the preparation of many pure enantiomers.[3–5, 100, 101] The ability of lipases to catalyze highly selective reactions under mild reaction conditions as well as their wide availability, high efficiency and stability, make them very attractive and valuable catalysts for the industrial production of fine chemicals.[4, 7–9]

One of the most efficient lipases to catalyze the enantioselective acylation of racemic amines and secondary alcohols is *Candida antarctica* lipase B (CalB).[68, 102–105] This enzyme also ex-

hibits high enantioselectivity in the hydrolysis of several racemic secondary esters, yielding enantiomerically pure esters and alcohols.[106–108] CalB therefore seems very suitable for the preparation of enantiopure chiral amino alcohols (compounds which contain both an amino and a hydroxy group available to be acylated) like Propranolol. Indeed, CalB-catalyzed acylation or hydrolysis reactions have been succesfully used to prepare this type of compounds.[109, 110]

1.2.2. *Candida antarctica* lipase B

From the yeast *Candida antarctica* two lipases have been isolated: CalA and CalB.[111, 112] CalB has a molecular weight of 33 kDa and is composed of 317 amino acid residues. The three dimensional structures of CalB and CalB complexes have been determined through the use of X-ray crystallography (PDB codes: 1TCA, 1TCB, 1TCC, 1LBS and 1LBT).[113, 114] All lipases contain a characteristic catalytic triad. In CalB this catalytic triad consists of Ser105, His224 and Asp187 (see figure 1.4, free CalB).

The generally accepted mechanism for reactions catalyzed by lipases involves two steps (figure 1.4 shows CalB-catalyzed acetylation):

1. The acylation step, where (in CalB) Ser105 attacks the carbonyl carbon atom of an ester substrate (the acyl donor) to form a tetrahedral intermediate (TI). The tetrahedral intermediate releases an alcohol (propranolol), yielding the acyl-enzyme (AcCalB).

2. The deacylation step, where the acyl-enzyme reacts with H_2O (or other nucleophiles, such as alcohols, amines or peroxides[5]) to give a carboxylic acid and the free enzyme.

Similar to the CalB-catalyzed acetylation of (*R,S*)-propranolol (figure 1.4), the CalB-catalyzed hydrolysis of propranolol esters involves two steps. In the hydrolysis reaction, the enantioselectivity is determined by the acylation step, when the ester acyl group is transferred to the catalytic Ser105 yielding a reactive acyl enzyme (AcCalB) and propranolol (see figure 3.5).

Acylation and deacylation proceed via an initial noncovalent enzyme–substrate complex (Michaelis complex, MCC) and an oxyanionic TI. The tetrahedral intermediates are stabilized by hydrogen-bonding interactions with the OH and NH groups of the so-called oxyanion hole of the enzyme (composed of residues Thr40 and Gln106 in CalB, see figures 1.4 and 3.5).[115, 116]

The active site of CalB is buried in the core of its structure and consists of two hydrophobic pockets. When viewed with the catalytic triad oriented with Asp left and Ser right, above the catalytic triad there is a large hydrophobic pocket and below the catalytic triad there is a medium-size pocket (see figure 3.6a). On the left side of he large pocket we find Ile189
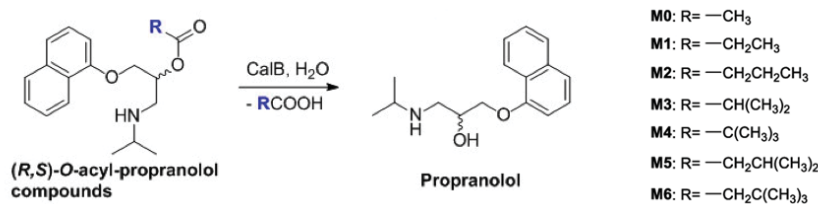
Figure 1.5.: What is the effect of the acyl moiety on the enantioselectivity of CalB catalyzed hydrolysis of propranolol esters?

and Val190, on the right there are Val154 and Leu144 and Leu140 is located at the top. The medium-size pocket is surrounded by Trp104 and the Leu278–Ala287 helix (helix α10).

1.2.3. Hydrolysis reactions of ester compounds

Computational methods have been used to study the acetylation of propranolol catalyzed by CalB.[91, 117–119] It was found that using vinyl acetate as acyl donor leaves enough space in the catalytic cavity of CalB for either *R*- or *S*-propranolol to fit in a reactive conformation, which would partly account for the observed moderate enantioselectivity.[91, 117–119] This raises the question of whether using an acyl donor with a longer or bulkier acyl chain could help to introduce steric differences. This would potentially enable CalB to distinguish the propranolol enantiomers and thus improve the enantioselectivity, as previously observed for other acylation reactions catalyzed by lipases.[120–123]

Up to now, the effect of the reaction conditions on the resolution of propranolol via CalB-catalyed acylation has not been studied in detail.[89, 91] Moreover, lipase-catalyzed hydrolysis reactions of ester compounds, which constitute an alternative to acylation reactions in the preparation of enantiomerically pure alcohols, have not been explored. This means there are many open questions about the potential of CalB to efficiently resolve the racemic mixture of propranolol. In the work discussed in this thesis we have therefore investigated the enantioselectivity of the CalB-catalyzed hydrolysis of propranolol esters, with a focus on the effect of the acyl moiety on the enantioselectivity (see figure 1.5).

1.3. Computational chemistry

For both the conversion of GGPP to taxadiene and the hydrolysis of propranolol esters, insights into the enzymatic reaction mechanism could open up the possibility for rational design of enzymes, mutations or better substrates. To acquire this kind of mechanistical information,

computational methods are a useful tool.[20, 38, 63, 66, 67, 72, 73, 124–129] They can provide data at an atomistic level about the individual catalytic steps. There are several computational methods, with distinct advantages and disadvantages, that can be used to study enzymatic reactions.

Quantum Mechanics (QM) Gas-phase QM calculations can be used in smaller systems (up to 1000 atoms, but for practical applications we are typically talking about up to around 100 atoms), to study the interactions leading to a particular geometry and the forming and breaking of bonds.[130] QM methods can be used to analyze the inherent reactivity of the substrate,[38, 44, 131] but quickly become expensive due to their scaling behavior. They are therefore usually applied to the substrate in isolation or in the presence of small models of functional groups present in enzyme active sites.[38, 44, 118, 131, 132]

Molecular Mechanics (MM) For larger systems MM can be used in conjunction with Molecular Dynamics (MD) to study enzyme motions and to sample configurational space. This is important to fully understand the origin of the catalytic properties of enzymes, because enzymatic reactions are known to involve multiple configurations of enzyme-substrate complexes.[118, 133–136] However, MM methods do not allow for bond breaking and forming.

Multiscale Modelling In combined quantum mechanics/molecular mechanics (QM/MM) calculations, the electronically important part of the system is treated at the QM level, allowing chemical reactions to happen, while the remainder of the system is treated at the MM level, which reduces the cost of the calculations significantly. This allows the study of catalytic pathways in enzymes, taking into account the effects originating from the active site, the remainder of the protein and the solvent environment. It is the interplay of these environmental effects that gives rise to enzyme efficiency and (stereo-) selectivity.[137–141]

1.4. Outline

In the following chapter, the applied computational methods will be discussed in more depth. Thereafter, in chapter 3, the results from three computational studies are summarized: two on the pathways from GGPP to taxadiene and one on the hydrolysis of propranolol esters. These studies provide detailed insight into the role of the TXS and CalB enzymes in catalysis. This will be followed by a brief conclusion and outlook in chapter 4.

2. Theoretical Background

This chapter gives a slightly more elaborate discussion of the methods used in chapter 3 than is provided in the Introduction. The aim is to present a short overview of quantum chemistry with a focus on QM/MM methods. Interested readers are referred to the cited publications and quantum chemistry textbooks for more details.[130, 142–144]

2.1. Molecular Mechanics

Molecular mechanics (MM) is a method to study large systems (like macromolecules in a solvent surrounding) with high efficiency. Using classical (Newtonian) mechanics and force fields (see section 2.1.1) the potential energy (U) of the system is defined as a function of the nuclear coordinates. This way one can find the molecular structure of a (local) energy minimum by energy minimization, using an algorithm (like steepest descent or conjugate gradient) to minimize the potential energy of the system. When MM is used to model the evolution of a system over time, this is called molecular dynamics (MD, see section 2.1.2).

2.1.1. Forcefields

To calculate the potential energy of a system at the MM level, energy functions and a set of potential parameters for that particular system are needed, which together are called the force field. Force fields are formulated empirically and are parameterized against a set of reference properties using experimental and/or theoretical data. Electronic degrees of freedom are typically neglected and atoms (the nucleus and electrons together) are described as point particles.

In this thesis, the CHARMM27 force field[145, 146] is employed, which includes bonded and non-bonded interactions between these point particles. The potential energy function of the CHARMM27 force field is a sum of sums of different types of contributions, see equation 2.1. The bonded energy terms comprise bond stretching, angle bending, dihedral and improper dihedral energy terms. The last two terms of equation 2.1 are the non-bonded energy terms: the Coulomb interactions, which model the electrostatic interactions between the point charges

at each atomic position; and the attractive and repulsive van der Waals (vdW) interactions represented by Lennard-Jones terms.

$$
\begin{aligned}
U = & \sum_{\text{bonds}} K_r(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB}(r^{1-3} - r_0^{1-3})^2 \\
& + \sum_{\text{dihedrals}} K_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{\text{impropers}} K_\omega(\omega - \omega_0)^2 + \sum_{\text{residues}} u_{CMAP}(\Phi, \Psi) \\
& + \sum_{\text{coulomb}} \frac{q_i q_j}{r_{ij}} + \sum_{\text{VdW}} \varepsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^6 \right]
\end{aligned}
\tag{2.1}
$$

Here the force constants K; the equilibrium geometry values (indicated with the subscript 0, e.g. r_0, θ_0, etc); the partial charges q_i and q_j; and the Lennard-Jones parameters (ε, r^{min}) are force field parameters that are taken from the literature or need to be parameterized. The bond length and bond angle terms of the bonded interactions are treated in harmonic approximation. In equation 2.1 r denotes the distance between atoms i and j, θ is a bond angle, and φ is a dihedral angle. The torsional energy is represented by a cosine function, with periodicity n and the phase angle δ. The improper torsion potential energy term involves the improper angle ω that measures the deviation from planarity, and may thus serve to maintain planarity.

Most terms in equation 2.1 are standard terms of molecular mechanics force fields, with the exception of the Urey-Bradley term and the CMAP term. The Urey-Bradley term is a constraint on the distance between atoms 1 and 3 of a bond angle. The CMAP term improves the conformational properties of the protein backbone, being a cross-term between the Φ and Ψ backbone dihedral angles.

In addition to a parameter file, MM calculations need a topology file in which molecular fragments are defined. In proteins these fragments are taken to be the amino acids that are their basic building blocks. The residue topology file contains entries for all amino acids and other residues like TIP3P water, assigning a unique type to atoms based on their chemical surrounding. The information about these fragments includes which atoms constitute the group, the connectivity of atoms within the group (including bonds and angles), and properties of the group like charge information.

2.1.2. Molecular Dynamics

Energy minimization provides a static image of the molecular structure, e.g. of an (local) energy minimum, while MD simulations show the temporal evolution of a system; the latter represent the main application of MM.

In an MD simulation, the atoms are assigned initial positions (e.g., from experimental or theoretical data) and velocities (e.g., from a randomized Gaussian distribution). Then, the force field is used to calculate the force on each atom. By numerical integration of Newton's equations of motion, the position and velocity of the atoms in the simulation is calculated for consecutive time steps. This makes it possible to study the behaviour of a system or the interplay between systems (e.g., solute and solvent interplay) as a function of time, for instance to sample configuration space.

Unfortunately, to study reactions where covalent bonds are broken and/or formed, electronic effects need to be considered, so MM does not suffice. For that we need to go beyond classical mechanics and enter the realm of Quantum Mechanics.

2.2. Quantum Mechanics

Quantum mechanics (QM) applied to chemical systems is called quantum chemistry. Chemical problems may be solved by combining experiment and quantum chemistry, but the research in this thesis is purely theoretical in nature. We perform calculations of ground-state properties of (macro)molecules utilizing computers. For this purpose the Schrödinger equation (section 2.2.1) has to be solved as accurately as possible.[130] This can be done using different QM methods which can be roughly divided into three families; ab initio methods (section 2.2.3), density functional theory (section 2.2.4) and semi-empirical methods (section 2.2.5).[130, 142]

2.2.1. Schrödinger equation

By solving the Schrödinger equation, the wave function (Ψ), energies and other properties of a system can be obtained. The time-independent Schrödinger equation has the following form:

$$\mathscr{H}\Psi = E\Psi(\mathbf{r}) \tag{2.2}$$

where Ψ is the wavefunction of the system; E denotes the associated eigenvalues and $\mathscr{H}$ the Hamilton operator equal to:

$$\mathscr{H} = \frac{-\hbar^2}{2m}\nabla^2 + \mathscr{V} \tag{2.3}$$

Equation 2.3 shows that the Hamiltonian is made up of kinetic (T) and potential energy ($\mathscr{V}$) terms. Solving equation 2.2 yields different solutions which correspond to stationary states of the system, where the *ground state* is the one with the lowest energy.

Only for one-electron or model systems (e.g. particle in a box, H atom, He$^+$ atom and Li^{2+} atom) can an exact solution to the Schrödinger equation be obtained.[143] However, by applying a number of approximations and assumptions it becomes possible to get approximate solutions for larger systems. One fundamental approximation is the Born-Oppenheimer (BO) approximation.[147]

Born-Oppenheimer approximation   In atoms and molecules, the electrons are significantly more mobile than the nuclei, because the mass of the nuclei and that of the electrons differ significantly while the forces on them are similar. The BO approximation makes use of this distinction by assuming that the electrons respond instantaneously to the motion of the nucleus so the nuclei look fixed to the electrons.[147]

As a consequence, it is possible to separate the wave function of a molecule into an electronic and a nuclear part. The Hamiltonian can be broken up likewise, so the Schrödinger equation can be solved using an electronic Hamiltonian that neglects the kinetic energy term for the nuclei. This yields the electronic wave function in the field of fixed nuclei. By repeating this for different nuclear positions the potential energy surface (PES) can be constructed. A PES describes the energy of the molecule in a given electronic state as a function of the nuclear coordinates.

Molecular orbital theory   In a second approximation, the N-electron wave function Ψ, is represented in terms of molecular orbitals: $\phi_1, \phi_2, ..., \phi_i$. This is called molecular orbital (MO) theory. The simplest way to construct Ψ as a combination of MO's is the Hartree product, but the resulting wave function is not adequate for fermions, like electrons, since it is not antisymmetric with regard to the exchange of electrons. By contrast, a Slater determinant built from orbitals is antisymmetric and satisfies the Pauli exclusion principle.

In open shell calculations it is assumed that each MO only holds one electron. However, most calculations use doubly occupied orbitals with two electrons of opposite spin (closed shell).

2.2.2. Basis sets

The molecular orbitals ϕ_i, can be approximated as a linear combination of atomic orbitals, χ_i (LCAO). These AOs are called basis functions and together they form the basis set.

An individual MO is defined as:

$$\phi_i = \sum_{\mu=1}^{N} c_{\mu i} \chi_\mu \tag{2.4}$$

where $c_{\mu i}$ are the LCAO coefficients and $\chi_1 ... \chi_N$ are the basis functions, chosen to be normalized. Gaussian-type functions are commonly used as basis functions.

A minimum basis set utilizes only a single basis function for each orbital on each atom. The flexibility of the basis can be increased by employing two or more functions of the same kind for each orbital (double-ζ or multiple-ζ basis sets). By adding functions with one extra node, polarization effects can be better represented (polarized basis sets, e.g. with a p-function added to valence s orbitals, etc). By including very shallow (diffuse) basis functions long range interactions can be better described, giving rise to extended basis sets. The larger the basis set, the more accurate results can be obtained.

2.2.3. *Ab initio* methods

Ab initio quantumchemical methods try to solve the electronic Schrödinger equation with only physical constants as input, i.e. *ab initio*. The simplest such a method is the Hartree-Fock method (HF).

Hartree-Fock In the Hartree-Fock (HF) method the wave function is approximated by a single Slater determinant. HF theory is also called mean field theory, because of the assumption that each electron only feels the average potential of the other electrons.

According to the variational principle, the lowest-energy Slater determinant is the closest we can get to the true wavefunction with that *ansatz*. The solution of the Hartree-Fock equations yields the variationally best orbitals:

$$F\phi_i = \epsilon_i \phi_i \tag{2.5}$$

Here F is the one-electron Fock operator generated; ϕ_i denotes the one-electron wave functions (*i.e.* the HF-MOs); and ϵ_i is the orbital energy.

As discussed before (section 2.2.2) the one-electron wave functions can be represented in

LCAO approximation. The introduction of an atomic orbital basis set transforms the Hartree-Fock equations into the Roothaan equations:

$$\mathbf{FC} = \mathbf{SC}\epsilon \qquad\qquad (2.6)$$

where $\mathbf{F}$ is the Fock matrix (depending on the coefficients C), $\mathbf{S}$ is the overlap matrix between the atomic orbitals, $\mathbf{C}$ is a matrix of coefficients and ϵ is a diagonal matrix of the orbital energies ϵ_i.

Through the orbitals the input for both the HF equations and the Roothaan equations depends on their solution. This means that both equations have to be solved by first guessing some initial orbitals, then solving the equations to get new orbitals and using these as new initial guesses. This way the orbitals are optimized in an iterative manner, until the change of the total electronic energy becomes smaller than a predefined threshold. This is why Hartree-Fock is also called a self-consistent field (SCF) method.

Post-Hartree-Fock　　Other *ab initio* methods are second-order Moller-Plesset perturbation theory (MP2),[148] and coupled cluster theory.[149] MP2 adds electron correlation effects and improves on the HF method by using perturbation theory to second order: $E \approx E_{HF} + E_{MP2}$. The coupled cluster approach accounts more accurately for electron correlation by using an exponential expansion of the wave function ($|\Psi\rangle = e^T|\Phi_0\rangle$). A CCSD(T) calculation (coupled cluster with single, double, and perturbative triple excitations) with a large basis set is the current "gold standard" of ground-state theoretical chemistry.

2.2.4. Density functional theory

In 1964, Hohenberg and Kohn published two theorems which demonstrated the existence of a unique functional that exactly determines the electron density $\rho(\mathbf{r})$, which governs ground-state properties of a many-electron system. The first theorem proves that the electron density determines the Hamiltonian ($\mathscr{H}$) and the wave function (Ψ), while the second theorem provides a method for its calculation.[150] Density functional theory (DFT) aims at finding the electron density for the electronic ground state of a system.

Kohn-Sham　　To find the ground-state energy, the Kohn-Sham approach requires solving a set of one-electron HF-like equestions, called the Kohn-Sham equations. The idea is to solve them for a system of non-interacting particles that, moving in an external potential, generates the same density as the real (interacting) system. The Kohn-Sham equations are

typically represented as:

$$\frac{\hbar^2}{2m}\nabla^2\Psi_j(\mathbf{r}) + V_{\text{eff}}(\mathbf{r})\phi_j(\mathbf{r}) = \varepsilon_j\phi_j(\mathbf{r}) \tag{2.7}$$

where $V_{\text{eff}}(\mathbf{r})$ is the effective potential giving rise to the same density as a real system and ε_j is the orbital energy of the Kohn-Sham orbital ϕ_j. The orbitals ϕ_j in turn yield the density by $\rho(\mathbf{r}) = \sum \left|\phi_j(\mathbf{r})\right|^2$

In the Kohn-Sham approach, the electronic energy can be written as:

$$E[\rho] = T_s[\rho] + \int \rho(\mathbf{r})v_{\text{ext}}(\mathbf{r})d\mathbf{r} + J[\rho] + E_{XC}[\rho] \tag{2.8}$$

where $T_s[\rho]$ is the kinetic energy of the system of non-interacting particles, v_{ext} is the potential acting on the interacting system and $J[\rho]$ is the electron-electron repulsion term (classical Coulomb interaction). The first part $(T_s[\rho] + \int \rho(\mathbf{r})v_{\text{ext}}(\mathbf{r})d\mathbf{r} + J[\rho])$ corresponds to the classical energy of the charge distribution. The $E_{XC}[\rho]$ term in equation 2.8 is the exchange-correlation energy, which accounts for the remaining non-classical terms in the energy. In practice, this exchange-correlation energy must be approximated.

Exchange-Correlation functionals   The $E_{XC}[\rho]$ term is usually divided into an *exchange* and a *correlation* part:

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho] \tag{2.9}$$

Both right-side components can be *local functionals* (depending only on the electron density ρ), and *gradient-corrected functionals* (depending on both ρ and its gradient, $\nabla\rho$).

One way to approximate the exchange-correlation energy is the local-density approximation (LDA). Here the exchange-correlation energy is approximated by assuming that the E_{XC} per electron at a point $\mathbf{r}$ in a gas is equal to the E_{XC} of a homogeneous electron gas (HEG), as long as the HEG has the same density in point $\mathbf{r}$ as the real gas. LDA functionals are mainly used in physics, but are generally not accurate enough for chemical purposes. They are, however, important for more advanced functionals.
The generalized gradient approximation (GGA) improves on the LDA by introducing a dependence of the exchange-correlation functional on the local gradient of the electron density $(\nabla\rho)$, generally improving molecular geometries and ground-state energies.[151–154]

The inclusion of HF exchange gives rise to so called hybrid functionals. A well known hybrid functional, formulated by Becke, is called the Becke-style three-parameter functional.[155] It

includes a mixture of the exact exchange energy from HF exchange, DFT exchange and DFT correlation:

$$E_{XC}^{B3LYP} = E_X^{LDA} + c_0(E_X^{HF} - E_X^{LDA}) + c_X \Delta E_X^{B88} + E_C^{VWN3} + c_C(E_C^{LYP} - E_C^{VWN3}) \qquad (2.10)$$

Here three parameters allow different admixtures of Hartree-Fock, LDA, Becke, Vosko-Wilk-Nusair and Lee-Yang-Parr contributions to be used. Different hybrid functionals can be constructed, either by varying the parameters or changing the component functionals.

2.2.5. Semiempirical methods

A third family of QM methods are semiempirical methods. They are not used in this thesis. In contrast to *ab initio* methods, these methods rely on empirical data and careful parameterization. They are based on the Hartree-Fock formalism, but employ a number of simplifying approximations.[156] This does make them very fast, and if the studied system is (sufficiently) alike to the systems used for parameterization, their accuracy can be quite good.

Since the enzymes studied here were too large to be treated entirely at the QM level, we employed a hybrid QM and MM method.

2.3. Quantum Mechanics/Molecular Mechanics

As the name implies, in combined quantum mechanics/molecular mechanics (QM/MM) calculations, part of the system is described at the QM level, while the remainder of the system is treated at the MM level. In computational chemistry, there is always a tradeoff between accuracy and efficiency. Treatment of the full system at the QM level would be accurate but costly, while treatment at the MM level would be fast, but would not allow for the study of chemical reactions.

Especially the study of enzymatic reactions is a good application of the QM/MM method, since the chemical reaction usually happens in a small part of the enzyme, the active site. And although the remainder of the system is important, it often suffices to include the environmental effects at the MM level. By treating the electronically important part (the active site) of the system at the QM level, chemical enzyme reactions can be studied with good accuracy and efficiency.

In the following sections we will discuss some aspects of the method, namely how to combine the energy contributions of the QM and MM part (section 2.3.1); how to handle the interaction of the QM and MM regions through non-bonded interactions (section 2.3.2); how to deal with

cutting through a covalent bond when defining the QM and MM regions (section 2.3.3) and how to deal with long-range interactions in a finite-size QM/MM system (section 2.3.4).

2.3.1. Subtractive or additive energy computation scheme

To compute the energy of the total system, the energy contributions from the QM and MM parts need to be combined. This can be done with a subtractive or an additive energy computation scheme.

The subtractive scheme can be represented as follows:

$$E = E_{MM_{total system}} - E_{MM_{QM region}} + E_{QM} \tag{2.11}$$

In words, this translates to first calculating the energy of the entire system using the MM forcefield, then removing the part associated with the QM region and replacing it with the QM energy of the QM region. In this interpolation scheme, the QM/MM interaction is approximated at the MM level, which is not very accurate and questionable if no appropriate force field parameters are available for the QM region.

Generally more accurate than the subtractive scheme is the additive scheme, which is used in this thesis. For the additive scheme, the QM region does not need to be parametrised, because the QM and MM region are calculated separately:

$$E = E_{QM_{QM region}} - E_{MM_{MM region}} + E_{QM/MM} \tag{2.12}$$

The interaction of the QM and MM region through non-bonded interactions ($E_{QM/MM}$) is handled by an embedding scheme.

2.3.2. Embedding

The simplest form of embedding describes the interactions between the QM and MM regions at the classical (MM) level. This is called mechanical embedding. Though steric effects are considered this way, the electrostatic interactions with the MM system have no effect on the wave function or electron density of the QM region.

To better account for the electrostatic QM/MM interactions, in the electrostatic embedding scheme,[157] MM point charges are added to the QM Hamiltonian. The van-der-Waals interactions are still treated at the MM level. This is the embedding scheme used throughout this work. To allow for polarization of the MM region (not used in this work) non-self-
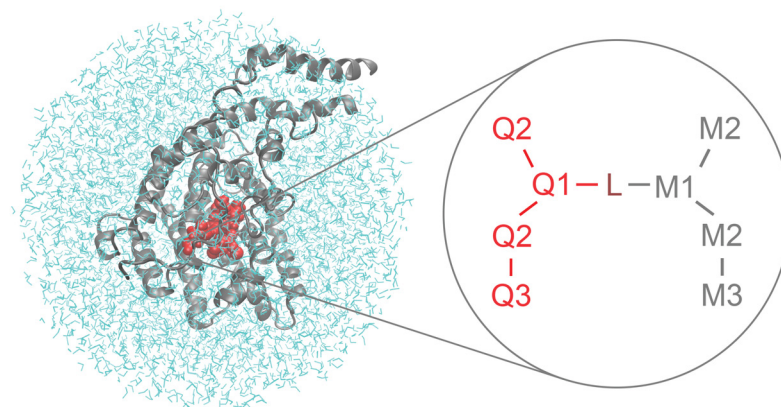
Figure 2.1.: Example of a typical devision of a QM/MM system into QM and MM region for an enzymatic reaction. In red: the QM region consisting of relevant protein residues, the substrate and possibly some specific ions and water molecules. The MM region consists of the remainder of the enzyme shown in gray and the solvent (water) shown in blue.

consistent[157] or self-consistent[158] polarizable embedding is required.

2.3.3. Link atom approach with charge shift scheme

It is often not possible to define a QM and MM region without cutting through a covalent bond, especially in enzymatic reactions where active site residues are involved in the reaction. This makes the treatment of the QM/MM boundary more difficult.

There are several ways to decently terminate the QM and MM region, while maintaining the intramolecular forces and bond(s) across the boundary. Unfortunately there is no 100% accurate, universally agreed upon method, which makes comparison of results between different programs more difficult.

Cutting a covalent bond generally causes open valencies in the QM region. To deal with this, in this thesis we use a strategy called the link atom approach. This approach introduces a link atom, usually a hydrogen atom, to saturate the (cut bond in the) QM region (see figure 2.1). To retain the bond between Q1 and M1 an MM bond stretch potential is added. The link atom introduces extra degrees of freedom, which can be removed by making the coordinates of the link atom dependent on the Q1 and M1 coordinates.

Because of the nearby M1 point charge the Q1-L bond can become over-polarized. By setting the charge on M1 to zero, this problem can be avoided. In the charge shift scheme[159] the

M1 charge is then shifted evenly onto the M2 atoms (see figure 2.1), to maintain the overall charge. By placing pairs of point charges in the proximity of M2, the M1-M2 dipole moment can be reproduced.

2.3.4. Boundary conditions and long-range interactions

For computational reasons, finite model systems are normally used in QM/MM calculations. To properly describe long-range electrostatic effects one can introduce periodic boundary conditions (PBC)[160] or solvent boundary potentials (SBP).[161] For the PBC approach, a simulation box is defined that is large enough to cover (most of) the enzyme to give a realistic representation of the enzyme environment. This box is then infinitely periodically repeated in all three directions, so that a molecule that leaves the box on one side will enter the box on the opposite side. In some of the calculations in this thesis we use PBC, which is very suitable for MM and QM/MM calculations.[162]

2.4. Computational treatment of chemical reactions

The methods discussed above provide energies at given geometries. We next discuss how to use this information to describe chemical reactions and to connect our findings to experiments. This can be done by a static PES-based approach (section 2.4.1) or by free energy methods (section 2.4.2).

Potential energy surface   The energy of a collection of atoms (molecule) as a function of their position (molecular geometry) constitutes the potential energy surface. The PES has the same dimensionality as the number of geometric degrees of freedom of the molecule (3N-6), where N is the number of atoms. However, in many applications, it suffices to focus on one distiguished coordinate (or a few of them). For example, chemical reactions often involve mainly one coordinate, the Reaction Coordinate (RC). A PES along the RC is called a potential energy curve or reaction profile.

A reaction profile starts in the reactant state and ends in the product state. In between, there is the transition state (TS), which is characterized as being a maximum in energy in one direction (the reaction coordinate) and a minimum in all other directions. It is also known as a first-order saddle point.

2.4.1. Static treatment of reactions

A typical reaction coordinate for a chemical reaction is a distance between atoms. Other RCs are linear combinations of distances, bond angles or dihedral angles. A scan along a RC can be constructed by performing a series of constrained geometry optimizations, in which the RC is kept fixed at suitably chosen values.

A more elaborate alternative is to add a harmonic potential to the total energy along the RC. For a bond-distance RC this potential would look like:

$$f_{\text{restraint}} = K_r (r - r_0)^2 \qquad (2.13)$$

with K_r the force constant, r the distance and r_0 the reference distance. This potential is identical to the bond term used in the force field (equation 2.1). The reaction profile is then obtained by a series of unconstrained geometry optimizations at suitably chosen values of r_0.

Transition state   The highest point of such a scan can be used as a guess for the structure of the transition state. This TS structure can then be further refined using e.g. a partitioned rational-function optimization (P-RFO) algorithm.[163, 164] The P-RFO algorithm uses an approximate or an exact Hessian. The gradient and the Hessian matrix can be calculated for each point on the PES as the first and second derivatives of the energy, respectively. The negative of the gradient gives the vector of forces on the atoms in the molecule. Diagonalization of the mass-weighted Hessian matrix yields eigenvalues from which the vibrational frequencies can be calculated.

IRC-like calculations   To find the reactant and product geometries connected to the TS structure, an intrinsic reaction coordinate (IRC) calculation can be performed.[165, 166] The IRC is defined as the steepest-descent pathway in mass-weighted coordinates from the TS of a reaction to its reactants and products.[167] It can be determined by a series of constrained optimizations: steps with a fixed step size are taken along the negative gradient in a mass-weighted Cartesian coordinate system, while all other coordinates orthogonal to the negative gradient are optimized.

Since this procedure is expensive in a QM/MM setting, in this work IRC-like calculations are performed. In this approximate IRC procedure, a fraction of the normal mode eigenvector corresponding to the imaginary frequency of the transition state (see next paragraph) is added to (or subtracted from) the structure of the transition state. The new structure is subjected to a careful unconstrained geometry optimization (with small step sizes) and the resulting

structure is visually inspected to confirm that it is the reactant or product.

Characterization of stationary points   The gradient and the Hessian can be used to analyze the nature of a structure obtained in a geometry optimization. Minima and transition states are stationary points, at which the gradient (i.e. the forces) must be zero. The Hessian is also known as the force constant matrix, since the eigenvectors of the mass-weighted Hessian in Cartesian coordinates correspond to the vibrational normal modes.[168] Using normal mode analysis, a structure can be characterized as a minimum or a transition state. For a minimum, all of the eigenvalues of the Hessian are positive, so the vibrational frequencies are real. For a TS, the Hessian has one (and only one) negative eigenvalue which corresponds to one imaginary vibrational frequency.

2.4.2. Free energy methods

Though free energy methods are not used in this work, they are recognized in section 3.2.3 as a possible way to resolve the discrepancy between some of our calculated results and experimental findings. This section will therefore briefly address the basic ideas underlying free energy perturbation (FEP)[169] and umbrella sampling,[170] two widely used free energy methods.

Free energy perturbation   To compute the difference in free energy when going from state A to state B with FEP, one runs a normal MD simulation for state A and also computes the energy for state B at each newly generated configuration. Computing the average over the simulation for state A, the free energy difference can then be calculated with the following equation:

$$\Delta F(A \rightarrow B) = -k_B T \ln \left\langle \exp \left(-\frac{E_B - E_A}{k_B T} \right) \right\rangle_A \qquad (2.14)$$

Umbrella sampling   When sampling a system with free energy barriers along the RC, an unbiased MD simulation will not provide a useful distribution function, because most of the simulation time will be spent in the low-energy regions, while the barrier regions (and thus the TSs) are rarely visited. In umbrella sampling a more even distribution is produced by introducing a biasing potential that forces the MD to visit configurations in the barrier regions that would be practically inaccessible otherwise. Then, the unbiased distribution can be retrieved from the biased one. One method of unbiasing or analyzing a series of umbrella sampling simulations is the weighted histogram analysis method (WHAM).[171]

3. Summaries of publications

3.1. Molecular Dynamics Study of Taxadiene Synthase Catalysis

Andrés M. Escorcia, **Jeaphianne P. M. van Rijn**, Gui-Juan Cheng, Patrick Schrepfer, Thomas B. Bruck and Walter Thiel

I carried out the MD calculations and performed the subsequent analysis for two of the four cations. I performed the analysis of the bond distances of all cations. I contributed to the writing of the manuscript.[66]

In this study we performed QM/MM MD simulations of enzyme-substrate complexes between the SHM of TXS and cations **C**, **F**, **D1** and **E** (hereafter denoted as ^{TXS}cation complexes, figure 1.1).[66] In order to study the dynamic behavior, we ran multiple 1 ns MD simulations of the ^{TXS}cation complexes in explicit water using different initial velocity distributions. We examined the enzyme structure and the cation conformations as observed in the MD simulations and addressed the hypothesis of cation tumbling. Based on average values of relevant interatomic distances[117,172] we discussed the reaction pathway and proposed deprotonation paths to the (side) products.

Computational details   For each of the ^{TXS}cation complexes three 1 ns MD simulations with different initial velocity distributions were carried out, to enhance sampling.[117,118,173–176] The chosen MD setup is analogous to that of previous studies reported in the literature.[117,118,177,178] All MD simulations were done with the CHARMM software package (version 35b2).[179] We applied a QM/MM MD approach[63,117,118] at 300 K using a Langevin thermostat.[180] The QM region (i.e., the carbocation) was treated by the self-consistent charge density functional

tight binding (SCC-DFTB) method.[181] The MM region (comprised of TXS, PPi, the catalytic Mg^{2+} ions, and the solvation layer) was described by the CHARMM27 force field.[145, 146] The QM/MM electrostatic interactions were calculated as implemented for the SCC-DFTB method in the CHARMM program.[182] The group-based extended electrostatics approach was used within the MM region;[183] the electrostatic interactions between particles closer than 14 Å were treated by the conventional pairwise additive scheme, while the interactions at larger distance were approximated by a computationally cheaper multipole approach. A quartic spherical boundary potential was applied to the water molecules to keep the shape of the water sphere and to prevent evaporation of outer water molecules.[179] All distances involving hydrogen atoms were constrained by SHAKE.[184] To assess the chances of the side chains of active-site residues to be involved in the deprotonation of the carbocations, we used the PROPKA module of the PDB2PQR server to predict their pKa values.[185–187]

3.1.1. TXS structure

Throughout all MD simulations we found that the structure of closed TXS is well conserved and the active site remains shielded from the bulk solvent, excluding solvent participation in the deprotonation of the carbocations. To keep the active site enclosed, residues Y89, H90, D92, Y841, and N845 play a key role by forming hydrogen bonds between the flexible regions (see appendix A, figure 4 and supporting information table S3). For example, Y89 forms hydrogen bonds with R580 (helix C) and/or D614 (helix D), and/or a water-mediated hydrogen bond with PPi. In addition, Y841 forms a water-mediated hydrogen bond with R580 and π-stacking interactions with F834 (helix J) and a hydrogen bond with S587 (helix C). The interactions of Y89 and Y841 pull the NTRC segment and the J-K loop toward the catalytic cavity of TXS .

The reaction cavity that harbors the carbocations is rigid, as indicated by RMSF values of 0.3-0.8 Å for most of the active-site residues. This rigidity appears to be due to two main factors:

1. The polar residues retain very strong and often multiple interactions with their surroundings during the MD simulations.

2. There is not enough space for the residues with bulky side chains to move freely, because their chains are oriented toward the inside of the catalytic cavity

The entrance of the active site is enriched in polar residues and harbors the catalytic Mg^{2+} ions and PPi. Therefore, some water molecules remain inside the active site after its closure, enabling the aforementioned water-mediated hydrogen bonds. The presence of water molecules in the active site is consistent with the reported crystal structures of class I terpene synthases

in closed form, in which the Mg^{2+} ions and PPi are surrounded by a significant number (13-17) of water molecules (see e.g. PDB entries 2OA6, 1N23, 1N20, and 4OKZ).[62, 69, 188]

3.1.2. Reaction pathway

As stated in the introduction, the class I terpenoid cyclase TXS initializes the reaction by GGPP ionization in which the pyrophosphate anion (PPi) is released. The electrostatic interactions between PPi, which is considered to remain in the active site during the entire cyclization cascade, and the carbocations are expected to have a large influence on the thermodynamics of the reaction.[20, 62, 63, 68, 69, 71] The distances between PPi:O1 and the location of the plus charge in the different cation complexes were used as an indication of the stability of the ^{TXS}C-^{TXS}F complexes in the HT-QM mechanism (see figure 1.1). Based on the PPi:O1–plus charge distances the ^{TXS}F complex is expected to be more stable than ^{TXS}C. The large PPi:O1-$D1$:C7(+) distance implies that the electrostatic interactions between PPi and the cationic center contribute little to the stabilization of $^{TXS}D1$, certainly less than in the case of ^{TXS}C and ^{TXS}F. ^{TXS}E is expected to be favored thermodynamically over the other complexes.

The average values of the relevant interatomic distances for the intramolecular proton transfer from cation C suggest that C will rearrange more easily to cation F than to cation $D1$ in the TXS environment, since the average ^{TXS}C:C2-H10 distance is 2.1 Å, while the average ^{TXS}C:C6-H10 distance is 2.5-2.6 Å (see figure 1.1). The same is found in the QM gas-phase calculations,[38] though calculations with the FHM indicate a preference for the direct pathway.[72, 73]

The A ring and the C ring The MD simulations of the $^{TXS}D1$ and ^{TXS}F complexes reveal a boat-like conformation of the A ring ($cation_{boat}$) in addition to the chair-like conformation ($cation_{chair}$) as found before in gas phase (see figure 1.1).[38] The cationic structure of $^{TXS}F_{chair}$ is equivalent to the HT-QM structure,[38] and is present over 60% of the MD simulation time, compared to less than 40% for $^{TXS}F_{boat}$. In one of the three MD runs $^{TXS}D1_{boat}$ is predominant (over 50% of the simulation time), while for the other runs the main observed conformation is $^{TXS}D1_{chair}$ (around 99%).

In the ^{TXS}E complex, the A ring of cation E has a boat-like conformation;[38, 45] however, the B and C rings are found in different conformations in the MD simulations. This results in three different conformers of the ^{TXS}E complex, which are labeled as E_{chair}, $E_{t\text{-}boat}$, and E_{B1} (see figure 3.1). E_{chair} is equivalent to the (starting) HT-QM structure.[38] The difference between $E_{t\text{-}boat}$ and E_{chair} is mainly in the C ring, which has a twist boat-like conformation in the former and a chair-like conformation in the latter (see gray indent in figure 1.1). E_{B1} differs
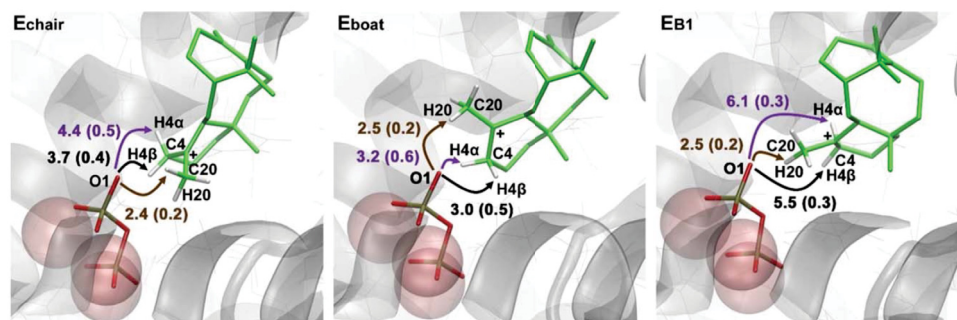
29

Figure 3.1.: Conformers of cation **E** identified in the MD simulations of the ^{TXS}**E** complex. Arrows show possible reaction paths for the deprotonation of cation **E** to **T** and **T1**. Average values for the relevant interatomic distances and the associated standard deviations (in parentheses) are given in Å. For each conformer, averages were calculated over all MD runs in which this conformer was encountered.

from **E**$_{chair}$ in the conformation of the B ring, as well as in the orientation of the C ring with respect to the other rings.

Carbocation tumbling The pyrophosphate moiety has been suggested as the most likely base in the biosynthetic pathway from GGPP to taxadine in TXS[20,43] as well as for other reactions in other terpene synthases.[63,74,76] Additionally, an arginine residue (R580) located in the active site of TXS was suggested to act as a base.[20]

In some ^{TXS}**cation** complexes, large interatomic distances (3.7-4.8 Å) are observed between the hydrogen atom expected to be abstracted from the carbocation and the base (either R580 or PPi). In those instances, the deprotonation is suggested to involve carbocation tumbling,[20] which is thought to be possible due to three factors:

1. The active-site volume is significantly larger than the volume of the substrate.[43] Moreover, the accessible space increases with increasing cyclization as the reaction progresses.

2. The electrostatic attraction between the cations and PPi can steer parts of the cations towards PPi.[20]

3. The π-interaction of cations **C**-**F** with the surrounding protein residues is reduced compared with cations **A**-**B**.[20]

Through the extensive sampling of the system, our MD simulations allow us to assess the hypothesis of carbocation tumbling. We found that the rigidity of the active-site cavity restricts the positional freedom of the carbocations. The promiscuity of TXS should therefore
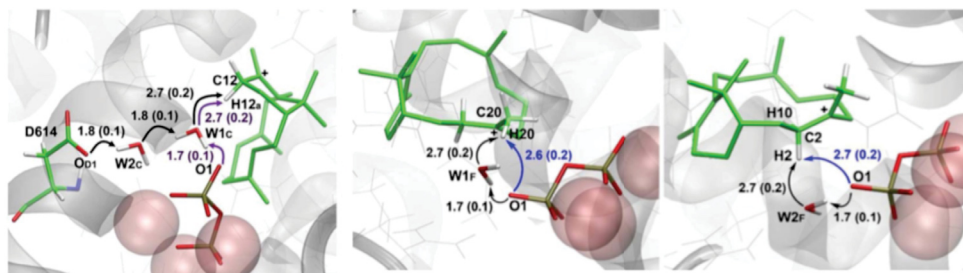
Figure 3.2.: a) Possible reaction paths for the water-assisted deprotonation of cation **C** (DPC) identified in the MD simulations of the $^{\mathrm{TXS}}$**C** complex: DPC1 (violet arrows) and DPC2 (black arrows). b and c) Possible reaction paths for the deprotonation of cation **F** (DPF) identified in the MD simulations of the $^{\mathrm{TXS}}$**F** complex: DPF1$_{C20}$ (left, blue arrow), DPF2$_{C20}$ (left, black arrows), DPF1$_{C2}$ (right, blue arrow), and DPF2$_{C2}$ (right, black arrows). Average values for the relevant interatomic distances and the associated standard deviations (in parentheses) are given in Å and correspond to one of the three MD runs. For more information see appendix A supporting information tables S5 and S7.

not be attributed to carbocation tumbling.

Deprotonation paths   As indicated in the introduction (section 1.1.2), several side products are observed (**V**, **V1**, **V2** and **T1**) in addition to the main product taxadiene (**T**).[20] These side products are expected to originate from the deprotonation of carbocation intermediates **C**, **F** and **E** (see figure 1.1).

The side product **V** is expected to result from deprotonation of the $^{\mathrm{TXS}}$**C** complex at **C**:C12. In the MD simulations we found three possible deprotonation pathways for cation **C**, with either PPi or an aspartate residue at helix D (D614) acting as the final proton acceptor. These pathways all include multiple proton transfer reactions[189–191] assisted by water bridges (see figure 3.2). Structures amenable to deprotonation reactions via one of these pathways are found to be present for 49-86% of the total simulation time.

Side products **V1** or **V2** stem from deprotonation of cation **F** at **F**:C20 and **F**:C2, respectively. In the MD simulation of the $^{\mathrm{TXS}}$**F** complex, the direct deprotonation of **F**:C20 to **V1** was found 30-70% of the MD simulation time; predominant over water-assisted deprotonation which was found for only 3-27% of the total simulation time. Water-mediated deprotonation of **F**:C2 to **V2** was encountered over 70% of the simulation time in the MD runs (see figure 3.2).

In the absence of carbocation tumbling, it does not seem feasible to deprotonate cation **D1** in the $^{\mathrm{TXS}}$**D1** complex.

31

Cation **E** can be deprotonated to the main product **T** (through deprotonation at **E**:C4) or the side product **T1** (through deprotonation at **E**:C20). For all three conformations of cation **E** found in the MD simulations of $^{\mathbf{TXS}}$**E** (**E**$_{\mathbf{chair}}$, **E**$_{\mathbf{t\text{-}boat}}$ and **E**$_{\mathbf{B1}}$), the only possible deprotonation mechanism is the direct proton transfer to PPi (see figure 3.1).

Deprotonation at C4 (to **T**) is more favorable in the **E**$_{\mathbf{t\text{-}boat}}$ conformation than in the other conformations. In **E**$_{\mathbf{t\text{-}boat}}$ deprotonation on the β face of C4 is expected to predominate over that on the α face, because the average PPi:O1-H4β distance is shorter than the PPi:O1-H4α distance (see figure 3.1). Elimination of H4β is also favored stereoelectronically in the twist-boat conformation of the C ring, all in agreement with labeling experiments using deuterated GGPP.[45, 46]

Meanwhile, all three conformers of cation **E** can easily undergo deprotonation at C20 (to form **T1**). Moreover, the computed distances between PPi and the hydrogen atoms bound to C4 and C20 of cation **E** indicate a preference for formation of **T1** over **T**, for all three conformers of $^{\mathbf{TXS}}$**E**. Therefore, based on the average distances, the predominance of **T** over **T1** in the product distribution cannot be rationalized.

Going beyond MD simulations of the $^{\mathbf{TXS}}$**C**-$^{\mathbf{TXS}}$**F** complexes and computing the energy profiles for the whole reaction scheme (including the deprotonation reactions leading to the side products **V**-**V2**) might resolve this problem. The results of such a study at the QM/MM level are reported in the next section.

3.2. QM/MM Study of Taxadiene Synthase Catalysis

Jeaphianne P. M. van Rijn, Andrés M. Escorcia and Walter Thiel

I carried out all calculations and performed the subsequent analysis. I wrote the manuscript and created all figures.[124]

In this study we went beyond MD simulations of the ^{TXS}**cation** complexes and performed QM/MM calculations to compute complete energy profiles for the conversion of GGPP to **T** in the TXS environment. Calculating the barriers of the proposed deprotonation pathways to the side products provided further insight into the promiscuity of TXS, and by investigating the conformational changes of the A and C rings we found a variant of the HT-QM mechanism.

Computational details For the QM/MM calculations of reaction profiles, we employed a static approach in analogy to previously reported studies of our group.[118, 177, 192, 193] The QM/MM calculations were done using the ChemShell program suite.[159, 194] The QM part of the system was treated at the DFT level (M06-2X[195]/TZVP[196]) using the Gaussian09 software.[197] The remainder of the system was treated at the MM level using the CHARMM27 force-field.[145, 146] The MM energies and gradients were computed by DL_POLY[198] as implemented in ChemShell. The interaction of the QM region with the point charges of the MM surrounding was handled by electrostatic embedding combined with the charge shift scheme.[157] After optimization of the ^{TXS}C snapshots, scans were carried out using a suitable reaction coordinate to obtain the pathway back to ^{TXS}GGPP, forward to ^{TXS}**T** and to all side products (figure 1.1). The highest point on a scan provided an initial guess for the corresponding transition state structure, which was subsequently optimized using the P-RFO algorithm for a core region (of 7 to 53 atoms) ,[163, 164] while treating the remaining non-frozen nuclei with the L-BFGS algorithm.[199, 200]

3.2.1. Snapshot selection

Representative snapshots were taken from the MD simulation of the ^{TXS}C complex[66] that served as starting points for the QM/MM calculations. In the MD simulations different water
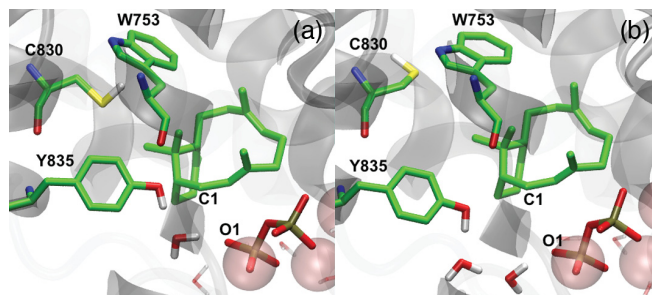
Figure 3.3.: a) ^{TXS}C:W1E1, with one water molecule between Y835 and PPi:O1; a PPi:O1-^{TXS}C:C1 distance of 3.3 Å due to NTRC orientation E1; and C830 interacting with W753, which plays a key role for the orientation of the C15(CH$_3$)$_2$ moiety of GGPP. b) ^{TXS}C:W2E2, with two water molecules between Y835 and PPi:O1; a PPi:O1-^{TXS}CC1 distance of 3.6 Å due to NTRC orientation E2; and C830 pointing away from W753. In both structures most hydrogen atoms are omitted for clarity. See in appendix B supporting information figure S2 for setup ^{TXS}C:W2E1 (E1 and two water molecules between Y835 and PPi:O1), setup ^{TXS}C:W1E2 (E2 and one water molecule between Y835 and PPi:O1) and setup ^{TXS}C:W1E2C (different orientation of cation **C**).

networks and enzyme conformations were observed, so for the selection of the snapshots we considered:

1. The occurrence (lifetime) of the particular configuration of the system in the MD simulations

2. The water network around PPi

3. The conformation and orientation of the cation

4. The conformation of the enzyme, in particular concerning the NTRC segment and the concomitant positioning of PPi in the active site (see in appendix B supporting information figure S1).

Based on these criteria, we selected four representative snapshots that had a long MD lifetime combined with two different water networks (indicated by the labels W1 and W2) and two different enzyme conformations (E1 and E2); these snapshots are labeled ^{TXS}C:W1E1, ^{TXS}C:W1E2, ^{TXS}C:W2E1 and ^{TXS}C:W2E2. A fifth snapshot, ^{TXS}C:W1E2C, has a different orientation of cation **C** in the binding pocket, which appeared for a shorter time in the MD simulation (see figure 3.3 and in appendix B supporting information figure S2).

W1 and W2 correspond to one or two water molecules mediating a hydrogen bond interaction between PPi and the adjacent residue Y835. E1 is characterized by an NTRC orientation that positions PPi close to the cation (the average PPi:O1-^{TXS}C:C1 distance in W1E1 and W2E1
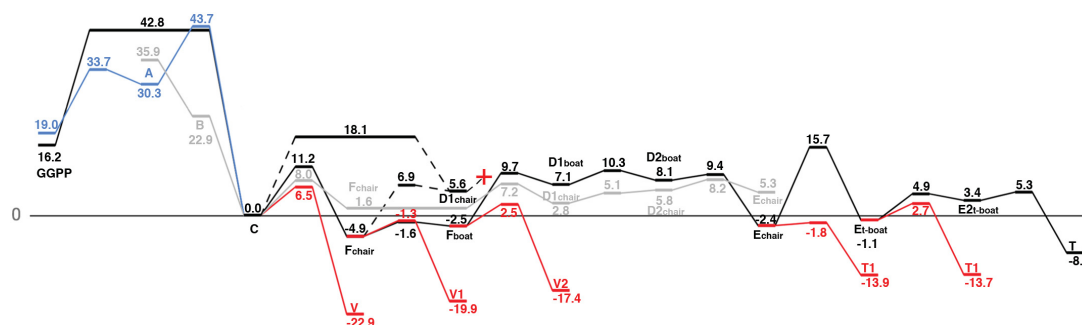
Figure 3.4.: QM(M06-2X/TZVP)/MM(CHARMM) reaction profile of the TXS-catalyzed cyclization of GGPP to taxadiene of setup ^{TXS}C:W1E1 (in black), GGPP-**C** for setup ^{TXS}C:W2E2 (in blue) and minor products (red). All energy values (kcal/mol) are relative to the ^{TXS}C complex. In gray: gas-phase QM SP energies (M06-2X/6-31+G(d,p), relative to cation **C**) on HT-QM structures[38] of cations **A**-**E**.

is 3.2 ± 0.1 Å) and the formation of an S-H··· π interaction between the thiol group of C830 and W753.[Footnote 1] In E2, the latter interaction is absent and the PPi:O1-^{TXS}C:C1 distance is 4.2 ± 0.8 Å (average of W1E2, W2E2, and W2E2C).

Six additional snapshots were taken from the MD simulations of the ^{TXS}F complexes[66] for computation of all pathways to the side products **V1** and **V2** discussed in section 3.1.2. For comparison to the energy profile obtained when starting from ^{TXS}C, a snapshot was taken from the MD simulations of ^{TXS}E,[66] snapshot W1E2E, which was propagated backward.

3.2.2. Reaction pathway

To determine the effect of the enzyme environment on the computed QM/MM energy profiles of the GGPP to **T** conversion, gas-phase QM calculations of the carbocation intermediates using density functional theory (DFT) serve as reference.[44, 63, 76, 77] These gas-phase QM energies are denoted with a superscript GQ (^{GQ}**cation**). Figure 3.4 shows the reaction profile for gas-phase QM as well as the QM/MM results for snapshot W1E1. The next paragraphs describe the five main differences between the QM/MM results and the gas phase HT-QM mechanism.

Pathway from ^{TXS}GGPP to ^{TXS}C In the TXS environment, the conformation of GGPP (^{TXS}GGPP) differs from that in the gas phase (^{GQ}GGPP)[38] mainly in the positioning of the

---

[1] In the MD simulation the thiol group of C830 flips from interacting with W753 (E1) to no interaction (E2). Because all snapshots with E2 are taken after the flip, the conformation of C830 is treated as a distinguishing feature between E1 and E2.

^{TXS}GGPP:C14=C15 double bond closer to the C1-O11 bond; and in the ^{TXS}GGPP:C10-C15 distance, which is shorter than that in the ^{GQ}GGPP structure (see figure 5a in appendix B).

The magnitude of these differences depends on the enzyme surrounding, which leads to a different pathway from GGPP to **C** depending on the NTRC position: when the NTRC is in E2 conformation, ^{TXS}GGPP first converts to ^{TXS}**A**, which then rearranges to ^{TXS}**C**, whereas for E1 a concerted pathway from GGPP to **C** is found (see figure 3.4).

Note that for all snapshots ^{TXS}**B** is not a minimum, which would explain the absence of side product **CM** in the wild-type product distribution (see figure 1.1)

Pathway from ^{TXS}C onwards The ^{TXS}**C** complex was found to rearrange more easily to ^{TXS}**F** than to ^{TXS}**D1** in agreement with previous findings (figure 3.4).[20,38,66] This result contrasts recent studies that utilize the FHM of the closed TXS conformation, where a slight preference for the direct proton transfer is found.[72,73]

While the conversion of ^{GQ}**C** to ^{GQ}**F** is endothermic, it is facilitated thermodynamically in the enzyme, because the stabilizing effect of PPi through electrostatic interactions is larger in ^{TXS}**F** than in ^{TXS}**C** (see appendix B, supporting information figure S3).

In our MD simulations we found that the A ring in ^{TXS}**F** can be in a chair- and boat-like conformation, as discussed in section 3.1.2. Both ^{TXS}**F$_{chair}$** and ^{TXS}**F$_{boat}$** are stable complexes, with ^{TXS}**F$_{boat}$** being 3.4 ± 1.3 kcal/mol less stable than ^{TXS}**F$_{chair}$**. This is consistent with the distribution observed in our MD simulations: ^{TXS}**F$_{chair}$** > 60% of the simulation time, versus ^{TXS}**F$_{boat}$** < 40%.[66] **F$_{boat}$** appears to only be stable in the enzyme, as it converts to **F$_{chair}$** in a gas phase QM optimization.

Pathway from ^{TXS}F to ^{TXS}D2 There is no clear preference for the conversion of ^{TXS}**F** to ^{TXS}**D1** with a chair-like or the boat-like conformation of the A ring. However, the subsequent formation of the productive conformer (^{TXS}**D2**, figure 1.1), through a conformational change of the 12-membered ring of ^{TXS}**D1** could only be found going from ^{TXS}**D1$_{boat}$** to ^{TXS}**D2$_{boat}$**, but not with the A ring in chair-like conformation. These findings suggest a conversion of the A ring to boat-like conformation earlier in the process, compared to the HT-QM mechanism where the A ring remains in chair-like conformation until cation **E** formation.[38]

Pathway from ^{TXS}D2$_{boat}$ to ^{TXS}E The conversion from ^{TXS}**D2$_{boat}$** to ^{TXS}**E$_{C\_chair}$** (A ring in boat- and C ring in chair-like conformation; see figure 1.1) has a low barrier (0.8 ± 0.6 kcal/mol) and a reaction energy of -12.6 ± 2.0 kcal/mol. The conversion of ^{TXS}GGPP to ^{TXS}**E$_{C\_chair}$** is exothermic by 17.3 ± 1.8 kcal/mol. Based on distances in previous computational work[20,66]

^{TXS}**E** was expected to be the energetic minimum of the pathway. However, ^{TXS}**F$_{chair}$** and ^{TXS}**F$_{boat}$** are more stable than the ^{TXS}**E** complex, because the stabilizing effect of PPi on these complexes is comparable (see appendix B, supporting information figure S3) while cation **F** is intrinsically more stable (see figure 3.4).[38] Nevertheless, ^{TXS}**E$_{C\_chair}$** is over 7 kcal/mol more stable than in the gas phase (see figure 3.4).

In the MD simulations of the ^{TXS}**E** complex, the C ring was found in both chair (^{TXS}**E$_{C\_chair}$**) and boat (^{TXS}**E$_{C\_boat}$**) conformations (see figure 3.1 and section 3.1.2).[66] Due to steric hindrance of C20 by PPi (see appendix B, supporting information figure S9), the barrier from ^{TXS}**E$_{C\_chair}$** to ^{TXS}**E$_{C\_boat}$** is 19.2 ± 1.5 kcal/mol for snapshots with the NTRC in E1, while snapshots with E2 have a much lower barrier of 9.3 ± 1.7 kcal/mol (see appendix B, supporting information tables S7, S8 and figure S9). A rotation of ^{TXS}**E$_{C\_boat}$** in the binding pocket leads to the novel ^{TXS}**E2$_{C\_boat}$** complex, for all but the two W2E2 snapshots, with barriers of 4-15 kcal/mol (see figure 3.4 and appendix B, figure 5b).

W1E2E   From the MD simulations of ^{TXS}**E**, an additional snapshot of ^{TXS}**E2$_{C\_boat}$** was taken, W1E2E (see appendix B, supporting information figure S9). Propagating the reaction backward from ^{TXS}**E2$_{C\_boat}$** to ^{TXS}GGPP yields an energy profile that is more in line with the FHM results[73] (see appendix B, supporting information table S9).

Deprotonation paths   We analyzed the expected preferred deprotonation pathways of the carbocations as found in the MD simulations (see section 3.1.2). Our calculations explain the promiscuity of the enzyme but not the product distribution. The barriers for the deprotonation paths for the W1E1 complex are shown in figure 3.4.

- The water-assisted deprotonation of ^{TXS}**C**:C12 by PPi:O1 to form ^{TXS}**V** was found to have a low barrier of 5.6 ± 3.0 kcal/mol and to be exothermic by -23.4 ± 2.4 kcal/mol.

- Formation of **V1** through direct deprotonation of ^{TXS}**F**:C20 by PPi:O1 has an average barrier of 4.4 ± 1.2 kcal/mol, and the process is exothermic by -12.2 ± 2.8 kcal/mol.

- For **V2** formation the direct deprotonation of ^{TXS}**F**:C2 by PPi:O1 (one snapshot) is barrierless, while the water-assisted deprotonation of ^{TXS}**F**:C2 by PPi:O1 has a barrier of 2.7 ± 2.6 kcal/mol. The reaction of **V2** formation is exothermic by -18.2 ± 2.0 kcal/mol.

- In ^{TXS}**E$_{C\_chair}$** and ^{TXS}**E$_{C\_boat}$** deprotonation of the α face of E:C4 by PPi:O4 (see figure 3.1) to yield **T** is feasible and favored over deprotonation of the β face by PPi:O4 or PPi:O1 (see appendix B, supporting information tables S8 and S10).
  After rotation to ^{TXS}**E2$_{C\_boat}$**, deprotonation on the β face (as found in experiments[45]) by PPi:O1 becomes favorable with a barrier of 2.7 ± 1.7 kcal/mol.

- Deprotonation of $^{TXS}E_{C\_chair}$:C20 and $^{TXS}E_{C\_boat}$:C20 by PPi:O1 to yield **T1** has barriers of 2-17 kcal/mol. The energy barrier for **T** formation from the $^{TXS}E2_{C\_boat}$ complex is similar to or lower than that of **T1** formation from the other ^{TXS}E complexes.

- Also for setup W1E2E (see appendix B, supporting information figure S10) the barrier to ^{TXS}T is 3.5 kcal/mol lower than the barrier to $^{TXS}T1$ (see appendix B, supporting information table S9). For W1E2E premature deprotonation by PPi:O1 to the side products seems unlikely (based on distances, see appendix B, supporting information table S12).

Curiously, these computed barriers and reaction energies indicate that the formation of some of the minor products is more facile than the formation of **T**, as can also be seen in figure 3.4.

3.2.3. Discussion

There are several possible explanations for the perceived preference for side product formation in our calculations.

Water molecules   The active-site water molecules in TXS influence the orientation of the cations and enable the low-barrier water-assisted deprotonation paths that can prematurely terminate the reaction. To our knowledge there is no experimental evidence on the number of water molecules remaining in the binding pocket of the enzyme after closing, though the involvement of water as a base in the catalysis of terpene synthase reactions (except for enzymes generating hydroxylated products) is an uncommon notion.[201] The crystal structures of terpene synthases show only a few water molecules to be trapped in the active site close to the carbocations, and they are usually well stabilized by their surroundings in positions where they are not well oriented to react.[56, 62, 201–203] However, the large active-site volume of TXS could mean that TXS is different from other terpene synthases.[43, 66]

Dynamic effects   Although the barriers and reaction energies for formation of the side products are favorable, there might be a higher number of TXS cation complexes leading to formation of taxadiene[66] than to formation of the side products, due to dynamical effects.[135, 136]

Product release   It has been shown for terpene synthases that though ionization of the diphosphate ester bond is the rate-limiting chemical step, product release is the rate-limiting step in the overall reaction.[76, 204, 205] Considering that the taxadiene complex is 1-15 kcal/mol

less stable then the other product complexes (see appendix B, supporting information tables S6-S7, S11-S12), taxadiene is expected to be released more easily, which may contribute to the higher yields of **T** observed experimentally.

Different enzyme states The two published models of TXS in the closed conformation, the SHM[20] and the more recent FHM,[72] differ appreciably and give rise to different energy profiles. Both models, however, might reflect reality. It may be possible that the FHM and SHM reflect states of the TXS catalysis with high and low population that favor formation of taxadiene and of the side products, respectively.

Enzyme model Our QM/MM results on TXS catalysis (as shown in figure 3.4 and appendix B supporting information table S6) deviate from recent QM/MM results obtained with the FHM.[73] Our pathway is less downhill from $^{\textbf{TXS}}$**A**, shows a preference for the two-step pathway to $^{\textbf{TXS}}$**D** via cation **F**, does not include $^{\textbf{TXS}}$**B**, and does include conformational changes of the A and C rings (figure 3.4). The presence of $^{\textbf{TXS}}$**F** in the pathway accounts for side products **V1** and **V2**, while the absence of $^{\textbf{TXS}}$**B** is in line with CM not being a side product of the TXS catalysis. The differences in the results from the present and previous[73] QM/MM studies may partly be due to the different methods employed for computing the reaction profiles (static QM/MM vs. free energy QM/MM calculations). We believe, however, that the differences between the underlying structural models as outlined in the introduction (section 1.1.3) are more important.

Substrate orientation The QM/MM reaction profiles for the five chosen snapshots give an internally consistent qualitative scenario for TXS catalysis but they also differ appreciably in a quantitative sense. Furthermore, the additional snapshot W1E2E taken from the MD simulations of $^{\textbf{TXS}}$**E2**$_{\textbf{C\_boat}}$ yields an energy profile that differs rather strongly from the others and is more in line with the FHM results;[73] we note that the orientation of the cation with respect to PPi differs in the W1E1 and W1E2E snapshots (see appendix B supporting information table S10). In general, it is very difficult to predict the correct bound state of the substrate, intermediates, and product in terpene synthases.[74, 76, 206] This mainly due to the absence of hydrogen bonds between the ligand and the enzyme that could serve as anchors to keep the former in place.[74] This may be especially challenging for TXS considering its very large active site.[43] Thus, though the structures of the $^{\textbf{TXS}}$**cation** complexes used in this study were obtained from a commonly employed docking[20] and MD procedure,[66] it could still be possible that relevant orientations of the cations have been missed.

The sensitivity of our static QM/MM results with regard to the chosen snapshot suggests that

it would seem worthwhile to perform QM/MM dynamics simulations with the SHM, which might be helpful to rationalize the product distribution of TXS.

3.3. Effects of the Acyl Moiety on CalB-catalyzed Hydrolysis of Propranolol Esters.

Gui-Juan Cheng, Andrés M. Escorcia, **Jeaphianne P. M. van Rijn** and Walter Thiel

I carried out part of the QM/MM calculations. I wrote the scripts to determine the productive Michaelis complexes and for analysis of the bond distances. I helped interpret the results and contributed to the writing of the manuscript.

For CalB catalyzed hydrolysis of (*R,S*)-propranolol ester, the enantioselectivity is determined by the acylation step, which is shown in figure 3.5. To understand how the structure of the acyl donor affects the binding of the propranolol ester with CalB and the enantioselectivity, in the present work we conducted a computational study of the acylation step to investigate the CalB-catalyzed hydrolysis of propranolol esters with a range of acyl donors of various structures (linear and branched) and different chain lengths. The acyl donors of **M0**, **M1** and **M2** have a linear chain which is gradually elongated from **M0** to **M2**. For esters **M3** to **M6**, the acyl donors have branched alkyl chains (see figure 3.5).

Computational method The study of the propranolol esters follows the same procedure in all cases: First the propranolol ester is docked against CalB to explore the possible binding modes (see section 3.3.1). From this the best conformers are chosen for post-docking optimization to take potential induced fit effects into account (see section 3.3.1). Then productive Michaelis complexes (MCCs) are selected for MD simulations to sample their conformations and to evaluate their dynamic behavior (section 3.3.2). From the MD runs, snapshots of productive MCCs are submitted to further QM/MM calculations (section 3.3.3).

The approach used in this study and the aspects considered during the analysis of the simulations, are analogous to those of the previous computational studies on the acetylation of propranolol catalyzed by CalB.[91, 117, 118, 207] A more detailed description of the methods is presented in the corresponding sections as outlined above.
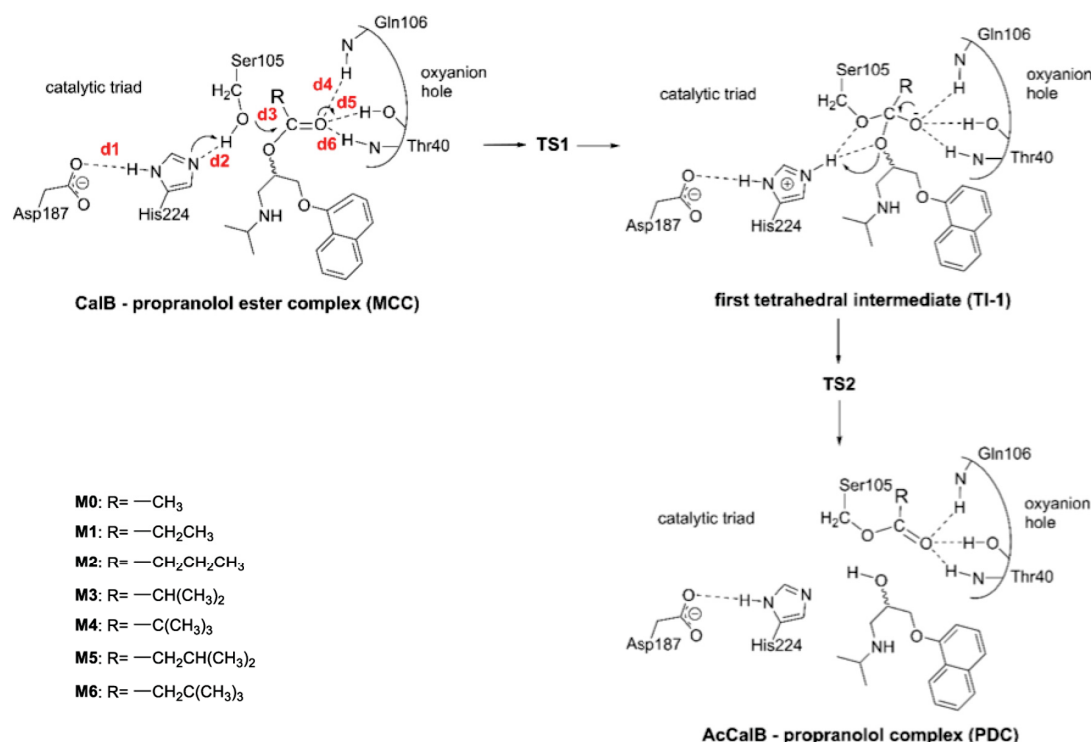
Figure 3.5.: Mechanism of the acylation step in the CalB-catalyzed hydrolysis of (*R,S*)-propranolol esters. Here the initial noncovalent ester-CalB complex (Michaelis complex; MCC) is transformed to the first tetrahedral intermediate (TI-1) which converts to an acylated CalB (AcCalB) product complex (PDC). TS1 and TS2 correspond to the relevant transition states. Hydrogen bonds important for the catalytic process are shown in dashed lines. **d1-d6** are the most relevant inter-atomic distances for the catalytic process concerning the MCC and are used here to qualitatively analyzed the relative reactivity of the MCCs of **M0-M6**, as obtained by combined docking and MD simulations (see the text for details). **M0-M6** are the compounds studied in this work.

3.3.1. Docking and post-docking optimization

Our simulations are based on the crystal structure of CalB with PDB code 1TCA.[113] This structure was protonated at pH 7 using PROPKA[186, 187] and solvated in a TIP3P water sphere with 40 Å radius. A representative structure of solvated CalB can be found in appendix C, figure 1. The solvated protein was subjected to a 1 ns NVT MD simulation at 300 K with the CHARMM software package (version 35b2),[179] using the CHARMM27 force field.[145, 146]

The overall structure of CalB was very stable throughout the MD simulation, with the residues of the catalytic triad being well organized for catalysis (forming relevant hydrogen bonds, see

figure 3.5) most of the simulation time. The dominant structure of CalB during the simulation was used as target to dock the propranolol esters **M0-M6**.

Docking

To construct the MCCs, the *R*- and *S*-enantiomers of seven propranolol esters, **M0-M6** (see figure 3.5), were non-covalently docked against CalB using the Autodock Vina software.[208, 209] Four docking runs were performed using different grid boxes and treating the protein either as rigid or semiflexible (i.e. with the side chains of a few residues treated as flexible). Duplicate CalB·ester complexes obtained from different runs were removed. For all complexes obtained from the docking procedure, the ability to be transformed into the AcCalB·propranolol product complex (PDC, see figure 3.5) was analyzed. The MCCs were considered capable of undergoing the hydrolysis reaction if they satisfied the following two criteria:

1. The distance between the hydroxyl oxygen of Ser105 and the carboxyl carbon of the propranolol ester (**d3**, figure 3.5) is shorter than or equal to 4.5 Å.[Footnote 2]

2. The carbonyl oxygen of the propranolol ester is oriented towards the oxyanion hole and forms at least one hydrogen bond with it (distances **d4, d5** and/or **d6** $\leq$ 3.0 Å, see figure 3.5).

The docking procedure with the seven propranolol esters against CalB resulted in 653 Michaelis complexes (MCCs) with affinity constants ranging from -2.4 to -7.5 kcal/mol. By considering the criteria above, a total of 36 MCCs were considered to be putative productive MCCs (i.e. complexes that may potentially lead to formation of the PDC). For propranolol esters **M0-M6** we found 9, 9, 8, 2, 2, 3, and 3 productive complexes, respectively.

Post-docking optimization

The selected MCCs were transferred to the CHARMM program (version 35b5)[179] for further optimization to take potential induced fit effects into account (i.e. subtle changes in the structure of the protein caused by the binding of the ligand).[211] Since no CHARMM force field parameters were available for the propranolol esters, a QM/MM approach was used.[91, 117, 118, 182] The QM region corresponds to the propranolol ester and was treated by the self-consistent charge-density functional tight binding (SCC-DFTB) method.[181] The MM

---

[2]The selection of a cutoff value of 4.5 Å for **d3** is based on previous computational studies on lipase reactions (including the acetylation of propranolol catalyzed by CalB) where a cutoff value of 4.0 or 4.5 Å for the distance between the reacting atoms has been successfully used to identify possible productive MCCs, providing results that are in good agreement with experiments.[80, 88, 93, 94, 210]

region (CalB and water) was described by the CHARMM27 force field.[145, 146] Three consecutive energy minimizations were performed, slowly lowering the constraints.[91] Only those complexes that satisfied the two criteria above and preserved the relevant hydrogen bonding network of the catalytic triad (distances **d1-d2** in figure 3.5) after optimization were still considered as putative productive MCCs.

A total of 15 complexes remained productive after optimization. For the *R*-(*S*-) enantiomer of esters **M0**-**M6** we found 3(3), 2(1), 3(0), 1(0), 0(0), 1(0), and 1(0) such MCCs, respectively. As can be seen, the propranolol esters with a branched acyl moiety (**M3**-**M6**) have less productive MCCs compared to the substrates with a linear acyl moiety (**M0**-**M2**). Especially for **M4** (with a bulky pivaloyl group) no productive MCCs remained after optimization. Moreover, **M3**, **M5** and **M6** have a longer distance between the carbonyl carbon and the catalytic serine (**d3** = 3.9-4.5 Å) than **M0**, **M1** and **M2** (**d3** = 3.4-3.8 Å). This suggests that substrates with a linear acyl moiety can better adapt to the binding site of CalB than those containing sterically demanding branched acyl groups.

Based on the relevant interatomic distances for the catalytic process (**d1-d6**) and the number of putative productive MCCs identified after optimization, we qualitatively assess the reactivity of the propranolol esters.[91, 117, 134, 135, 172, 212–214] The absence of productive MCCs for the *S*-enantiomer for **M3**, **M5** and **M6** suggests that the enantioselectivity of the reactions with these propranolol esters may be higher than for **M0**-**M2**. However, based on the interatomic distances and the number of productive MCCs, the reactivity of **M3**, **M5** and **M6** is expected to be lower than that of **M0**-**M2**. Therefore, only the productive MCCs of **M0**, **M1** and **M2** were selected for further MD simulations.

Binding modes   As indicated in the introduction (section 1.2.2), the binding pocket of CalB consists of a large and a medium-size hydrophobic pocket (see figure 3.6a).
As in previous studies on the CalB-catalyzed acylation reaction of propranolol,[91, 117, 118] the propranolol ester can fit into the binding pocket of CalB in two different binding modes; see figure 3.6. In binding mode I, the naphthoxy group of the propranolol ester is oriented towards the large hydrophobic pocket above the catalytic triad, while the isopropylamine side chain occupies the medium pocket below it and may extend toward the entrance. In binding mode II, the orientations of the naphthoxy and isopropylamine groups are reversed. Moreover, the acyl group may be oriented downward and bind to a small pocket deep inside CalB, which is composed of Thr42, Ser47, Trp104 and Leu278 (resulting in sub-binding modes Ia and IIa) or oriented upward close to Ile189 and Val190 (sub-binding modes Ib and IIb).

The acyl moiety affects the binding modes and reactivity. **M0** (with an acetyl group) can adopt both binding mode I and II, while for **M1** and **M2** (with a larger acyl group) all productive
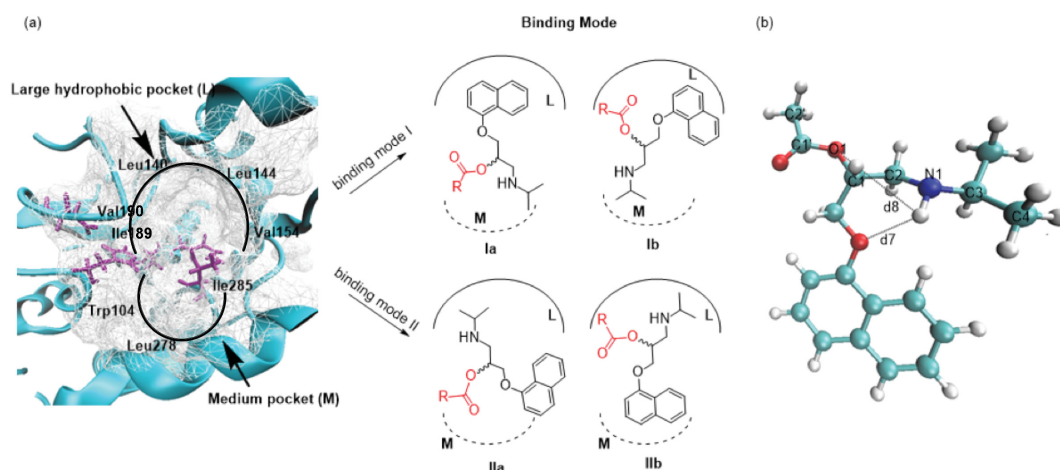
Figure 3.6.: a) The structure of CalB with the catalytic triad Asp-His-Ser (purple) oriented from left to right. Middle: Binding modes of propranolol ester in the CalB binding pocket. b) Hydrogen bonds used to distinguish between different configurations of **M0-M2**.

MCCs adopt binding mode I. **M3-M6** were also only observed in binding mode I. This is because the more space-demanding acyl groups of **M1-M6** drive the propranolol ester away from the catalytic site, exposing the naphthoxy group to the solvent in binding mode II. These conformations then quickly become non-productive during minimization.

3.3.2. Molecular dynamics study on Michaelis complexes of CalB-propranolol esters

Based on our qualitative assessment (see section 3.3.1) we expected some MCCs (**M0-M2**) to be more reactive than others (**M3-M6**). The optimized productive MCCs of the more reactive propranolol esters were selected for further analysis. The selected MCCs of **M0** (**M0R1**, **M0R2**, **M0R3**, **M0S1**, **M0S2** and **M0S3**), **M1** (**M1R1**, **M1R2** and **M1S2**) and **M2** (**M2R1**, **M2R2** and **M2R3**) are described along with their 3D structures in appendix C, figures 2-4, respectively. In addition to the 12 productive MCCs , two optimized MCCs identified as non-productive were also considered for MD simulations (**M1S1** and **M2S1**, see appendix C, figures 3-4), since a productive configuration was expected with better relaxation of the enzyme environment.

The selected MCCs were subjected to 1 ns NVT QM(SCC-DFTB)/MM(CHARMM27) MD simulations in explicit water (again using a TIP3P water sphere with 40 Å radius) to study their dynamic behavior.[117, 118, 172, 214] The QM region corresponds to the propranolol ester and

the MM region to CalB and the water. For each complex, two MD simulations with different initial velocity distributions were performed to enhance sampling.[66, 117, 118]

Analysis of the 28 MD simulations of the MCCs again focused on the identification of productive complexes. The enantioselectivity and the reactivity of the systems were qualitatively analyzed by considering the lifetime (i.e. time of occurrence during the MD simulations) of the productive MCCs, the relevant interatomic distances for the catalytic process (as indicated in figure 3.5), and the enzyme-substrate interactions.[117, 118] For all MD trajectories we also analyzed the stability of the catalytic triad (Asp-His-Ser hydrogen bonding, figure 3.5), the flexibility of the protein and the conformational diversity of the substrate.

Protein flexibility The structure of CalB is well conserved during the MD simulations and the hydrogen bonds between Asp187, His224 and Ser105 (**d1** and **d2**, see figure 3.5) are stable in all MD trajectories, indicating that the geometry of the catalytic triad is maintained during the simulations.
The flexibility of CalB was evaluated by considering the root mean square deviations (RMSD) of the protein backbone heavy atoms. In the MD runs the RMSD value increased gradually during the first 200 ps (heating and equilibration) after which it reached a plateau with a value of about 0.6 Å in all but one MD run (see appendix C, figure 5). Therefore, the MD trajectories from 200 to 1000 ps were used as productive trajectories for further analysis.

Substrate conformations The substrate, by contrast, exhibits more flexibility during the MD simulations and undergoes conformational changes along the trajectories. The average and deviation for distance **d3**, **d5** and **d6** as well as the life time for all productive conformations are recorded in appendix C, tables 1-3.

During the MD simulations, rotation along the N1-C2-C1-O1 dihedral angle is observed, leading to different hydrogen bond networks around the amino group (N1H, atom labels and hydrogen bonds **d7** and **d8** are defined in figure 3.6b). We distinguish four hydrogen bond networks (labeled **a** through **d**), where **a** corresponds to two intramolecular hydrogen bonds (**d7** and **d8** $\leq$ 3.0 Å), **b** to one hydrogen bond between the amino group and the ether group (**d7** $\leq$ 3.0 Å), **c** to one hydrogen bond between the amino group and the acetate group (**d8** $\leq$ 3.0 Å) and **d** to no intramolecular hydrogen bonds (**d7** and **d8** > 3.0 Å) (see figure 3.6b).
Not all hydrogen bond networks are observed in all MD simulations (see appendix C, tables 1-3). For example, for **R-M2** networks **a**, **b**, **c** and **d** are observed, while for **S-M2** only network **d** is found.

The propanoyl and butanoyl group of **M1** and **M2**, respectively, can adopt an extended or compact conformation (see appendix C, figures 6-7). The acyl group was observed to undergo

conformational interconversion between extended and compact structures during the MD simulations.

For **M1**, productive MCCs were identified for both the extended and compact structures of the propanoyl group (accessible by rotation around the C1'-C2' bond, see appendix C, figure 6). MCCs with a compact propanoyl group are found to approach Ser105 more closely (shorter **d3**) than MCCs with an extended propanoyl group (see appendix C, table 2, **M1R1** and **M1R1'**), since in the extended conformation the propyl group stretches toward to the catalytic site preventing the carbonyl group of the substrate access to Ser105.

For the butanoyl group of **M2**, rotation around the C1'-C2' bond positions the ethyl- and carbonyl group in an anti-conformation, while rotation around the C2'-C3' bond leads to an extended structure (see appendix C, figure 7). The CalB-**M2** complexes in which **M2** adopts an extended conformation (**M2R1** and **M2R3**) only stay productive for a few picoseconds. While when **M2** adopts a compact structure it could form favorable CH-π interactions with Trp104 and could stay in the binding pocket leading to longer lifetimes of productive MCCs (see appendix C, table 3).

Enatioselectivity   For **M0**, the carbonyl group of the propranolol ester is closer to the catalytic Ser105 in MCCs of **R-M0** (average values of **d3** between 3.72 and 4.08 Å) than in MCCs of **S-M0** (average values of **d3** between 4.12 and 4.40 Å). Additionally, two hydrogen bonds were formed between the carbonyl group of **M0** and the oxyanion hole in most MCCs of **R-M0**, while at most one hydrogen bond was observed for the MCCs of **S-M0** (see appendix C, table 1).

Similar observations can be made for **M1** and **M2**, where the shorter average **d3** distance and the increased number of hydrogen bonds with the oxyanion hole indicate that the *R*-enantiomer binds more favorably with the binding pocket of CalB than the *S*-enatiomer (see appendix C, tables 2 and 3).

Additionally, more productive MCCs were identified for the *R*-enantiomer than for the *S*-enantiomer for all propranolol esters, though this difference is most pronounced for **M2**. These findings suggest *R*-enantiopreference for the hydrolysis reaction of all three propranolol esters **M0-M2**.

Effects of acyl donor   Although a range of conformations of productive MCCs were identified for *R*- and *S*-propranolol esters **M0-M2**, the conformations we expect to be most favorable are similar for **M0**, **M1** and **M2** . In these conformations, the naphthoxy group binds to the large binding pocket and the isopropyl amino group occupies the medium pocket (Ia). These conformations are stabilized by CH-π interactions between the naphthoxy group and Ile189, Val154 and Thr138, CH-π interactions between the acyl group and Trp104, and hydrophobic

interactions between the isopropyl amino group and residues in the medium pocket.

Compared to **M0**, **M1** and **M2** may be quite enantioselective, due to the significant differences between their *R*- and *S*-enantiomers. For example, their *R*-enantiomers have a longer lifetime than *S*-**M1/M2** in most MD runs and the former have better (hydrogen bonding) interactions than the latter (see appendix C, tables 2 and 3). However, the reactivity of **M1** and **M2** is expected be low compared to **M0**, because: i) the binding process for **M0** is much more favorable (as indicated by a shorter **d3** distance and better hydrogen bonding interactions) and ii) for **M1** and **M2** the lifetime of productive MCCs is reduced for both *R*- and *S*-enantiomers and no reactive MCCs were identified for binding mode II (reinforcing the docking results). Therefore we focus on the most reactive propranolol ester, **M0**.

To get quantitative information about the enantioselectivity, representative snapshots of the systems were taken from the MD simulations of **M0** to be used as starting structures in subsequent QM/MM calculations of reaction profiles.[118, 215, 216]

3.3.3. QM/MM study on the hydrolysis reaction of (*R,S*)-O-acetyl-propranolol (M0)

The QM/MM calculations were performed with the ChemShell package,[159, 194] using Gaussian09[197] and DL-POLY[198] as QM and MM codes, respectively. For all systems, the QM region consists of the propranolol ester and the side chains of the residues of the catalytic triad (QM1 region). To test the importance of a particular water molecule for the catalysis, additional QM/MM calculations were performed including this water molecule in the QM region (QM2 region = QM1 region + water). Both QM regions have a total charge of -1.

In a previous QM/MM study of the acetylation of propranolol by CalB,[118] B3LYP/TZVP[155, 196] was validated against different density functionals with and without dispersion corrections, showing that the former is suitable for computation of reaction profiles of lipase reactions involving amino alcohols. Hence in all calculations, the QM region was treated at the B3LYP/TZVP level, while the rest of the system (MM region) was treated with the CHARMM27 force field.[145, 146] All atoms within 7 Å of the corresponding QM region (QM1 or QM2) were unconstrained during QM/MM optimization, whereas the positions of all other atoms were kept fixed. An electrostatic embedding scheme with charge shift correction was used for the treatment of the electrostatic interactions between the QM region and the surrounding partial charges of the MM region.[157, 159] Valencies at the covalent bonds crossing the QM/MM boundary ($C\alpha$-$C\beta$ bonds of the residues of the catalytic triad) were saturated using hydrogen link atoms.[158]

The computed reaction profiles involve five stationary points: MCC → TS1 → TI-1 → TS2 → PDC (see figure 3.5). We first optimized the snapshots of the MCCs taken from the MD

simulations. These optimized MCCs were used as starting points in potential energy surface (PES) scans along a suitably chosen reaction coordinate to reach the corresponding TI-1. The highest point on such a scan served as starting structure for a subsequent full optimization of TS1. Frequency calculations confirmed that the optimized TS1s were characterized by a single imaginary mode along the reaction coordinate. Intrinsic reaction coordinate like calculations and visual inspection ensured that a continuous pathway connected MCC and TI-1. Starting from TI-1, an analogous procedure was then applied to compute the TI-1 → PDC profile. Unless mentioned otherwise, all energies reported in this paper are given with respect to the energy of the MCC.

For a better understanding of the QM/MM reaction profiles, we performed single point QM calculations in the gas-phase (i.e. in absence of the MM region) for the QM region of all stationary points. The resulting $QM_{gas\text{-}phase}$ energies were subtracted from the corresponding original QM energies obtained in the QM/MM calculations (including the electrostatic interactions with the MM point charges) to quantify the stabilizing effect of the MM region on the QM region (QM_{elec} energy) along the reaction profiles.[118] Additionally, QM PES scans at the B3LYP/TZVP level for the rotation of the O-C ester bond of R-propranolol acetate were carried out to evaluate the influence of the orientation of the ester group on the reaction profiles (see appendix C, figure 8).

Hydrolysis reaction of *R*-M0   The QM/MM reaction profiles for the transformation of *R*-O-acetyl-propranolol (*R*-**M0**) to R-propranolol were computed for a total of five representative major conformations (**M0R1b**, **M0R1c**, **M0R1b'**, **M0R2d**, and **M0R3c**). **M0R2d** adopts binding mode IIb and the other four conformations are in binding mode Ia. Labels **a-d** are defined in section 3.3.2 and more details about the conformations can be found in appendix C, section 3. The QM(B3LYP/TZVP)/CHARMM energy profiles for the transformation of *R*-**M0** are shown in figure 3.7a.

In all reaction profiles the formation of the propranolol-CalB product complex (PDC) is the rate-determining step (TS2). The hydrolysis reactions of **M0R1b**, **M0R1c**, **M0R1b'**, and **M0R3c** are endothermic by 9.3-18.4 kcal/mol. The reaction profiles of **M0R1b** and **M0R1c** are fairly similar and kinetically favorable over those of **M0R1b'** and **M0R3c**, with overall barriers (MCC to TS2) for the former of 23-25 kcal/mol, while the latter have barriers of >30 kcal/mol (see figure 3.7a). The difference between the stationary points (TS1, TI and TS2) of **M0R1b'**/**M0R3c** and those of **M0R1b**/**M0R1c** primarily arise from higher MM energies in the former two (see appendix C, table 4). The higher MM energies of **M0R1b'** and **M0R3c** are mainly due to weakened interactions between the substrate and the surrounding residues during the course of the reaction (especially the CH-π interactions, see appendix C, table 5) and rearrangement of the binding pocket.
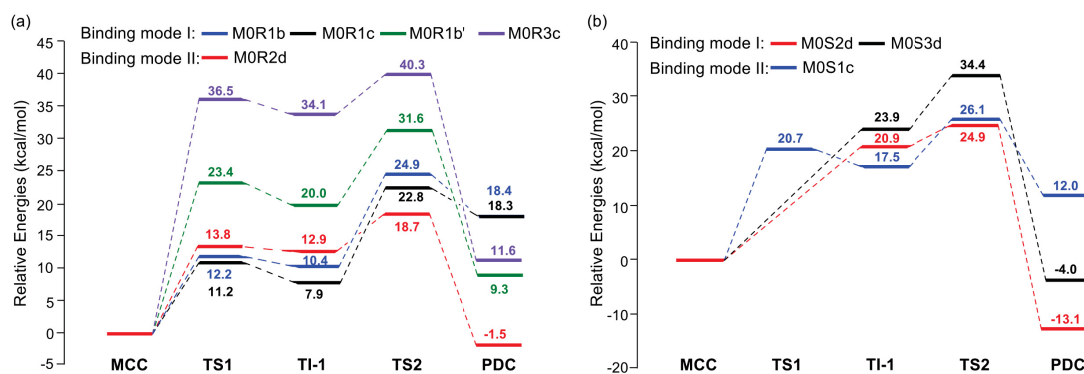
Figure 3.7.: QM(B3LYP/TZVP)/MM(CHARMM) energy profiles for the hydrolysis reaction of *R*- (a) and *S*-O-acetyl-propranolol (b) in binding modes I and II. Energies are given relative to the respective reactive complexes between CalB and propranolol esters (MCCs).

The hydrolysis reaction of **M0R2d** is exothermic by -1.5 kcal/mol and the overall reaction barrier is 18.7 kcal/mol, which is lower than for **M0R1c**. Therefore, the transformation of *R*-O-acetyl-propranolol via the **M0R2d** pathway is kinetically and thermodynamically more favorable than via the other pathways. The energy decomposition analysis (see appendix C, tables 4 and 6) shows that the relative QM energy of the PCD of **M0R2d** is over 9 kcal/mol lower than the other snapshots. Also $QM_{gas-phase}$ energies of TS2 and PDC in **M0R2d** are much lower than for the corresponding stationary points of other conformations (see appendix C, tables 4 and 6). The MM differences between **M0R2d** and the other conformations are less pronounced (see appendix C, table 4).

The reason for the favorable QM energies for **M0R2d**, is that in **M0R2d** the ester group converts from a *cis* conformation ($\omega_{C2'-C1'-O1-C1}$ = -27°) in the MCC to a *trans* conformation ($\omega_{C2'-C1'-O(Ser105)-C(Ser105)}$ = 164°) in the PDC. QM gas-phase calculations for the rotation of the ester group of **M0** indicate the *trans* conformer to be 6.2 kcal/mol more stable than the *cis* conformer (see appendix C, figure 8). In line with this result, the calculated $QM_{gas-phase}$ energy of the PDC is lower than that of the MCC by 6.4 kcal/mol for **M0R2d** (see appendix C, table 6).

By contrast, the ester groups of **M0R1b**, **M0R1c**, **M0R1b2** and **M0R3c** go from a *trans* conformer in the MCC to a (unfavorable) *cis* conformer in the PDC, leading to a positive $QM_{gas-phase}$ energy for the PDC for these setups (appendix C, table 6). This explains why the reaction is only exothermic for setup **M0R2d** while it is endothermic for the other setups.

Hydrolysis reaction of S-M0   The QM/MM reaction profiles for the transformation of *S*-O-acetyl-propranolol (**S-M0**) to *S*-propranolol were computed for a total of three representative major conformations (**M0S1c**, **M0S2d** and **M0S3d**). **M0S2d** and **M0S3d** adopt binding mode Ib and Ia, respectively. **M0S1c** adopts binding mode IIb. The QM/MM energy profiles are shown in figure 3.7b.

Similar to the transformation of **R-M0**, the formation of the propranolol product (TS2) is the rate-determining step in all reaction profiles. Attempts to locate TS1 for **M0S2d** and **M0S3d** failed due to the flat PES between TS1 and TI-1. **M0S1c** and **M0S2d** have comparable energies for TS2 of about 25 kcal/mol. However, the overall reaction of **M0S2d** is exothermic by -13.1 kcal/mol, while that of **M0S1c** is endothermic by 12.0 kcal/mol. The reaction of **M0S3d** has a much higher activation barrier (34.4 kcal/mol). Thus the reaction of **M0S2d** is both kinetically and thermodynamically more favorable than the reactions of **M0S3d** and **M0S1c**.

For all three energy profiles the hydrophobic interactions with the environment are important for stabilizing the system, as indicated by the negative relative MM energy contributions for all PDCs (see appendix C, table 7). The QM energy is the main factor favoring formation of the product through **M0S2d**.

For **M0S2d**, the conformation of the ester group changes from an unfavorable *cis* conformation in the MCC ($\omega_{C2'-C1'-O1-C1}$ = -15°) to a stable *trans* conformation in the PDC ($\omega_{C2'-C1'-O(Ser105)-C(Ser105)}$ = -176°) leading to a low QM energy (-8.8 kcal/mol) for the PDC of **M0S2d** (see appendix C, table 7). By contrast, the conformation of the ester group in the MCC and the PDC remains *trans* for **M0S1c** and *cis* for **M0S3d**.

Comparison between the reactions of R- and S-M0   The transformation of **R-** and **S-M0** via **M0R2d** and **M0S2d** are the most favorable reaction pathways for the formation of *R-* and *S*-propranolol, respectively. The reaction of **M0R2d** is kinetically more favorable than the reaction of **M0S2d** by 6.2 kcal/mol (see figure 3.7), which suggests a preference for the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol to generate *R*-propranolol, in agreement with the MD analysis. However, formation of *S*-propranolol is thermodynamically favorable.

For both **M0R2d** and **M0S2d**, the acetyl group occupies the large binding pocket and approaches Ser105 from the topside, leading to a *trans* ester in the PDC (see figure 3.8). In **M0R2d**, the naphthoxy group binds in the entrance of the medium pocket, allowing a water molecule to bind in the medium pocket, which assists the reaction through hydrogen bond interactions with the ester group. Since the isopropylamino group in **M0S2d** interacts with Trp104 and occupies the medium pocket, there is no space for a water molecule in the pocket

M0R2d-MCC   **M0R2d-TS2**   **M0R2d-PDC**
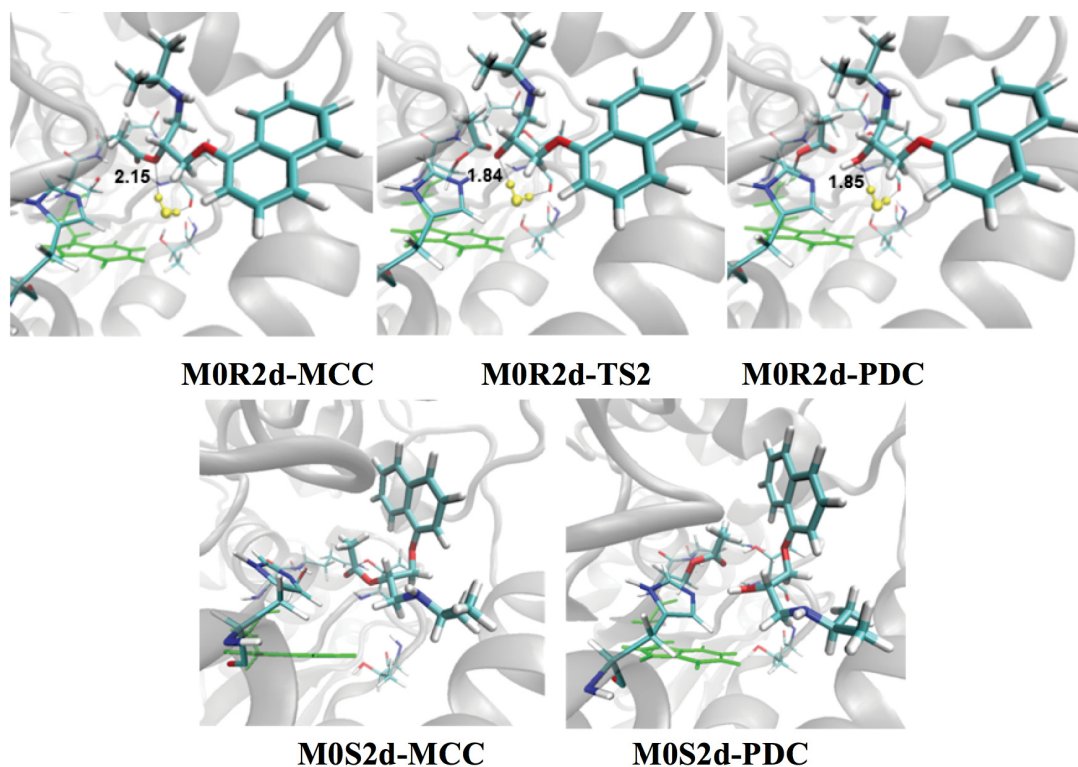
M0S2d-MCC   **M0S2d-PDC**

Figure 3.8.: Optimized structures of stationary points for the reaction of **M0R2d** and **M0S2d**.

itself in this case (see figure 3.8 and appendix C, figures 9 and 10).

The QM/MM energy difference between **M0S2d** and **M0R2d** mainly arises from the QM energy (see appendix C, tables 4 and 7), though single-point calculations demonstrate that the $QM_{\text{gas-phase}}$ energies for the stationary points of **M0S2d** and **M0R2d** are comparable (see appendix C, table 8).

During the reaction course the hydrogen bonding interactions between the substrate and oxyanion hole become stronger for both **M0R2d** and **M0S2d** (see appendix C, table 9), but for **M0R2d** there is the additional hydrogen bonding interaction between the water molecule in the medium binding pocket and the substrate (figure 3.8 and appendix C, figures 9 and 10). This water forms hydrogen bonds with Thr40 and the ester group of the substrate, stabilizing the developing charges at the oxygen atom of the substrate during the reaction course ($d_{\text{WAT:O}\cdots\text{M0:O2}}$ = 2.15 Å in the MCC reducing to 1.85 Å in the PDC (figure 3.8 and appendix C, figures 9 and 10 and table 9). Additional QM/MM calculations of **M0R2d** with the water molecule included in the QM region yield similar potential energies for the stationary points (see appendix C, table 9).

3.3.4. Summary and comparison to acylation reactions

The acetylation of propranolol catalyzed by CalB has been studied using computational methods.[91,117–119] Combined docking and molecular dynamics (MD) simulations were used to model the relevant enzyme-substrate/intermediate complexes,[91,117–119] while hybrid quantum mechanics/molecular mechanics (QM(B3LYP/TZVP)/MM(CHARMM)) calculations were used to compute the respective reaction profiles.[118] The focus was on the deacylation step of the reaction (see figure 1.4), which is responsible for the enantioselectivity in acylation reactions.[118]

The QM/MM calculations show that formation of O-acetyl-propranolol from *R*-propranolol is exothermic in binding modes I and II (by -2.6 and -8.8 kcal/mol, respectively), but only in binding mode I for *S*-propranolol (by -2.8 kcal/mol). In both enantiomers, formation of the second tetrahedral intermediate (TI-2) is found to be the rate-limiting step (TS1). TS1 for *S*-propranolol (binding mode I) is higher in energy than that of *R*-propranolol in both binding mode I and II (by 4.5 and 1.6 kcal/mol, respectively). Therefore the transformation of *S*-propranolol to O-acetyl-propranolol is expected to be slower, consistent with experimental results where an enantioselectivity is observed that corresponds to a free energy difference of 2.4 kcal/mol in favor of *R*-propranolol transformation.

In this work a combined approach of docking, molecular dynamics (MD) and hybrid quantum mechanics/molecular mechanics (QM(B3LYP/TZVP)/MM(CHARMM)) calculations was used. The CalB-catalyzed hydrolysis of propranolol esters was modeled with a range of acyl donors of various structures and different lengths (see figure 3.5) in order to understand how the structure of the acyl donor affects the binding of the propranolol ester with CalB, reactivity and enantioselectivity. This work focuses on the acylation step (figure 3.5), since for the CalB catalyzed hydrolysis of (*R,S*)-proprapanolol esters, the enantioselectivity is determined by this step after which propranolol is released.

Docking results suggest that acyl donors with branched alkyl chains are too sterically demanding to be reactive. Molecular simulations of the propranolol esters with linear chains suggest the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol will prefer to generate *R*-propranolol. Based on the MD simulations we expect the reactivity of **M0** to be higher than for **M1** and **M2**, since the lifetime of productive MCCs for both *R*- and *S*-**M1/M2** is lower than **M0** and no reactive MCCs were identified for binding mode II for **M1/M2**.
In agreement with MD results, the QM/MM calculations of the hydrolysis reaction of **M0** suggest an enantiomeric preference of the *R*-product. For the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol, the activation energy gap between the reaction of *R*- and *S*-**M0** is 6.2 kcal/mol. This gap is larger than the ones found for the acylation reaction (1.6 and 4.5 kcal/mol), indicating a potentially higher enantioselectivity for the hydrolysis reaction.

3.4. Further publications: contributions as co-author

Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases

Patrick Schrepfer, Alexander Buettner, Christian Goerner, Michael Hertel, **Jeaphianne P. M. van Rijn**, Frank Wallrapp, Wolfgang Eisenreich, Volker Sieber, Robert Kourist, and Thomas Bruck

Proceedings of the National Academy of Sciences, **2016**, 113, 958–967.

I analysed the data and helped interpret the results.

4. Conclusions and Outlook

In this thesis we focus on two important drugs: Taxol (an anticancer drug)[2,11,12] and Propranolol (a beta-adrenergic blocking agent used for treatment of cardiovascular disorders).[13–15]

For Taxol, the current commercial production processes heavily depend on the taxus plant and produce significant toxic waste streams, making them less environmentally sustainable and increasing the cost of taxol.[1,2,10,37] Propranolol is commercially available as racemic mixture, though the *S*-enantiomer is significantly more active than both the *R*-enantiomer and the racemic mixture,[15,79] while the racemic mixture has been shown to cause serious side effects.[80]

For both these pharmaceuticals, computations can provide valuable insights into the enzymatic reaction mechanism.[20,38,63,66,67,72,73,124–129] A more detailed understanding of the molecular basis of TXS promiscuity and the enantioselectivity of CalB-catalyzed hydrolysis, including knowledge about key enzyme residues, could open up the possibility for rational design of enzymes, mutations or better substrates to improve taxadiene formation and the propranolol enantioselectivity, respectively.

In this thesis MD and QM/MM calculations have been presented for two enzyme systems.

4.1. Taxadiene syntase catalysis

The first committed step in the biosynthetic pathway from GGPP to taxol is the conversion of GGPP into taxa-4(5),11(12)-diene (taxadiene, **T**), catalyzed by Taxadiene Syntase (TXS).[40] This step is the focus of the work presented in sections 3.1 and 3.2. There we aim to gain a more detailed understanding of TXS catalysis and provide new insights into the molecular basis of TXS promiscuity and the preference for taxadiene formation.

Four essential reaction intermediates in the conversion of GGPP into taxadiene are cations **C**, **F**, **D1** and **E** (see figure 1.1). Though no crystal structures of closed TXS complexed with GGPP or carbocation analogues have been reported, Schrepfer et. al. provide a homology model (the SHM) of TXS in the closed conformation.[20] We performed QM/MM MD simulations of enzyme-substrate complexes between the SHM of TXS and these cations (**^{TXS}cation**

complexes, figure 1.1 and section 3.1). These simulations provide insight into the promiscuity of TXS by identifying possible pathways to the four observed side products (**V**, **V1**, **V2**, and **T1**, figure 1.1). In agreement with a previous static study (employing docking and MM optimizations),[20] we find that the PPi anion, which is considered to remain in the active site during the entire cyclization cascade, plays a key role in TXS catalysis. The MD simulations do not support the hypothesis that carbocation tumbling enables deprotonation of all carbocations by PPi, since the cavity of the active site of TXS harboring the carbocations during the catalytic process is generally quite rigid and restricts the positional freedom of the carbocations.

Judging from the average values of the catalytically relevant interatomic distances in the MD simulations, in the $^{\mathbf{TXS}}$**C** complex, cation **C** is expected to be deprotonated to yield **V** via multiple proton transfer reactions assisted by water bridges, with either PPi or an aspartate residue at helix D (D614) acting as the final proton acceptor. In $^{\mathbf{TXS}}$**F**, cation **F** can be deprotonated either directly by PPi or by a water-assisted proton transfer to PPi to produce either **V1** or **V2**. In the absence of carbocation tumbling, it does not seem feasible to deprotonate cation **D1** in the $^{\mathbf{TXS}}$**D1** complex. Finally, an exclusive deprotonation of the carbocation by PPi is only possible in the $^{\mathbf{TXS}}$**E** complex. The MD simulations of the TXS cation complexes indicate that formation of **T** in the TXS environment via the HT-QM pathway requires a conformational change of the C ring of cation **E** from a chair-like to a boat-like conformation. In the resulting $^{\mathbf{TXS}}$**E** complex, cation **E** is in a much better orientation to be deprotonated by PPi at C4. This conformational change of the C ring correlates well with the results of labeling experiments using deuterated GGPP.[45, 46]

According to the QM/MM calculations (section 3.2), formation of minor products via water-assisted deprotonation of carbocation intermediates, as proposed in our MD study,[66] is energetically favorable, with barriers of 0-20 kcal/mol and reaction energies of -7 to -23 kcal/mol. We note, however, that water-assisted deprotonation is an uncommon notion for terpene synthases.[56]

The reaction pathway for the conversion of GGPP to **T** found in our calculations differs from previous proposals. It contains additional steps for conformational changes of the A ring in $^{\mathbf{TXS}}$**F** and the C ring in $^{\mathbf{TXS}}$**E** (see figure 1.1), as well as a rotation of cation **E** . Additionally, depending on the conformation of the enzyme, the $^{\mathbf{TXS}}$**A**→$^{\mathbf{TXS}}$**C** or the $^{\mathrm{TXS}}$GGPP→$^{\mathbf{TXS}}$**C** conversions may be concerted. The suggestion that $^{\mathbf{TXS}}$**B** is not a minimum on the PES contrasts other computational findings,[38, 73] but seems consistent with experimental evidence: stereochemical labeling experiments[45, 46] indicated the GGPP→**C** transformation could be a concerted process, and cembrene A is not observed in the product distribution.[20]

Our calculations identify routes to the observed side products and thus explain the promis-

cuity of the enzyme, but they do not reproduce the observed product distribution since the barriers to the side products are often lower than those to the next reaction step of the carbocation cascade. However, it is conceivable that dynamical effects disfavour the side reactions and that taxadiene is released more easily than the other products (see section 3.2.3).

The two published models of TXS in the closed conformation, the SHM[20] and the more recent FHM,[72] differ appreciably and give rise to different energy profiles. Both models, however, might reflect reality. The QM/MM free energy study with the FHM[73] provides an intriguing overall mechanistic scenario for TXS-catalyzed taxadiene formation, while our static QM/MM calculations offer detailed insight into the promiscuity of TXS. In this sense, the two studies are complementary to each other.

4.1.1. Outlook

In this thesis we discuss the work on the Taxadiene Synthase Mechanism of Schrepfer et. al,[20] Major and coworkers[72,73] as well as our own work.[66,124] These studies provide different models (the SHM and FHM) and different conclusions about TXS catalysis, regarding e.g. the carbocation minima, their conformation and relative stability, the deprotonating base, deprotonation paths and (expected) product distributions.

Though significantly different, the SHM and FHM both seem reasonable. Moreover, most of the different conclusions about TXS catalysis appear to be reasonable as well. Although it would seem that this situation arises from issues with the employed models, it is possible that the different models correspond to different biologically relevant states of TXS. Specifically, it could be that the FHM and SHM reflect states with high and low populations that favor formation of taxadiene and of the side products, respectively. If both models reflect some version of reality this means that several mechanistic issues remain unresolved. If we are dealing with a modeling problem, it might be that the currently employed methods are not adequate for a system like TXS.

The current studies underline the pronounced sensitivity of the TXS-catalyzed carbocation rearrangements to i) the enzyme environment and ii) differences of methodological nature. i) The sensitivity to the enzyme environment can be mainly attributed to strong electrostatic interactions that depend on the conformation of the cation, its positioning and orientation in the active-site, the active-site architecture and the presence of active-site water molecules. ii) The sensitivity to procedural details could point to the need to employ or develop more accurate homology modeling and docking methods, possibly even tailored to deal with terpene synthase enzymes.[217–221] The unusually large sensitivity of our static QM/MM results with regard to the chosen snapshot exemplifies the known limitations of the static QM/MM

approach to enzyme reactivity. It would therefore seem worthwhile to perform QM/MM dynamics simulations with the SHM, which might be helpful to rationalize the product distribution of TXS.

Though advances on the computational front might reduce differences between models or shed light on the cause of certain discrepancies between conclusions, it would ultimately take experiments to create the link between theory and the physical world. Regarding the Taxadiene Synthase Mechanism, this would mean experimental biochemists are challenged with the task of experimentally evaluating how valid both models are, with developing tests to tell them apart and with transforming the gained insights into ways to improve Taxol production.

4.2. *Candida antarctica* lipase B catalyzed hydrolysis

Lipase catalyzed hydrolysis reactions of ester compounds constitute an alternative to acylation reactions in the preparation of enantiomerically pure alcohols. In the study described in section 3.3, we set out to investigate the CalB-catalyzed hydrolysis of propranolol esters for a range of acyl donors of various structures and different lengths (see figure 3.5) to understand how the acyl donor affects the binding of the propranolol ester with CalB, the reactivity and the enantioselectivity. This work focuses on the acylation step (figure 3.5), since for CalB catalyzed hydrolysis of (*R,S*)-propranolol esters the enantioselectivity is determined by this step after which propranolol is released.

Docking results suggest that acyl donors with branched alkyl chains (**M3**-**M6**, figure 3.5) are too sterically demanding to be reactive. Subsequent molecular simulations of the propranolol esters with linear chains suggest the reactivity of esters with shorter chains (O-acetyl-propranolol, **M0**) to be high compared to propranolol esters with longer ones (O-propanoyl-propranolol, **M1** and O-butanoyl-propranolol, **M2**), based on the more favorable binding process for **M0**; the reduced lifetime of productive MCCs for both *R*- and *S*-**M1/M2** in the MD runs; and the fact that no reactive MCCs were identified in binding mode II for **M1** and **M2**. The MD simulations also suggest that the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol will prefer to generate *R*-propranolol.

In agreement with MD results, the QM/MM calculations of the hydrolysis reaction of **M0** suggest an enantiomeric preference for the *R*-product. The activation energy gap between the reaction of *R*- and *S*-**M0** is 6.2 kcal/mol, which is larger than that of the acylation reaction, indicating a potentially higher enantioselectivity for the hydrolysis reaction.

4.2.1. Outlook

Our study of the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol as described in section 3.3.3, leaves room for improvements on the computational front.

We found that the orientation of the ester group has a large influence on the reaction profiles (see appendix C, table 6-7, figure 8). For the most favorable pathway of both **R-M0** and **S-M0**, the acetyl group occupies the large binding pocket and approaches Ser105 from the topside, leading to a *trans* ester in the PDC. Meanwhile, the less favorable pathways of, for example, **R-M0** go from a *trans* conformer in the MCC to an unfavorable *cis* conformer in the PDC, leading to a positive $QM_{gas-phase}$ energy for the PDC for these setups (appendix C, table 6). QM gas-phase calculations for the rotation of the ester group of **M0** indicate the *trans* conformer to be 6.2 kcal/mol more stable than the *cis* conformer (see appendix C, figure 8). Since the orientation of the ester group has such a big effect on the final energy profile, it would seem worthwhile to take additional snapshots of both **R-** and **S-M0** to further study the effects of the conformation of the ester group on the reaction profiles.

Additional snapshots might also be taken to analyse the impact of a water molecule in the medium binding pocket of **R-M0**: energetically, the biggest difference between the most favorable pathways for **R-** and **S-M0** stems from the presence of this water molecule. During the reaction course the hydrogen bonding interactions between the substrate and oxyanion hole become stronger for the most favorable pathway of both **R-M0** and **S-M0** (see appendix C, table 9), but for **R-M0** there is an additional hydrogen bonding interaction between the water molecule in the medium binding pocket and the substrate (figure 3.8 and appendix C, figures 9 and 10). This water stabilizes the developing charges at the oxygen atom of the substrate during the reaction course (figure 3.8 and appendix C, figures 9 and 10 and table 9). Since the isopropylamino group in **S-M0** interacts with Trp104 and occupies the medium pocket, there is no space for a water molecule in the pocket itself in this case (see figure 3.8 and appendix C, figures 9 and 10). However, we currently did not analyse the lifetime of the water molecule in the medium binding pocket of **R-M0**. If the lifetime of this situation is short, the comparison with the most favorable pathway for **S-M0** should be with a snapshot for **R-M0** which would follow a similar path, but does not have the additional interaction with the water.

Instead of taking additional snapshots as described in the previous paragraphs, performing free energy calculations on the hydrolysis reaction would be preferable, as it would not only provide insight into the issues raised above, but also give additional data about the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol.

4. Conclusions and Outlook

On the other hand it would be interesting to turn to experiments. The study of CalB catalyzed hydrolysis of propranolol esters as discussed in chapter 3.3 indicates a potentially higher enantioselectivity for the hydrolysis reactions compared to CalB catalyzed acylation reactions. For the hydrolysis reaction of racemic (*R,S*)-O-acetyl-propranolol, the activation energy of the conversion of **R-M0** is 6.2 kcal/mol lower than that of **S-M0**, which is larger than the gaps found for the acylation reaction (1.6 and 4.5 kcal/mol). To see if this larger activation energy gap indeed translates to a higher enantioselectivity for the hydrolysis reactions, it would be intriguing to investigate this in a experimental setting.

Moreover, experiments could shed light on another issue. As described in section 3.3.3, while the reaction of **R-M0** is kinetically more favorable than the reaction of **S-M0**, formation of *S*-propranolol is thermodynamically favored by 11.6 kcal/mol. This would suggest that the final product in an experiment would depend on the reaction conditions, where short reaction times at low temperature would lead to the kinetic product *R*-propranolol, while a long enough reaction time at a temperature that is high enough to overcome all reaction barriers would lead to the thermodynamic product *S*-propranolol.

These additional studies could provide further molecular basis for improving the enantioselective synthesis of propranolol via lipase catalyzed reactions, through the rational design of either the reaction conditions or CalB.

Ultimately the goal of this research is to make enantiomerically pure propranolol, as well as more and cheaper taxol. Hopefully, the works in this thesis bring us a small step closer.

Bibliography

[1] S. Howat, B. Park, I. S. Oh, Y.-W. Jin, E.-K. Lee, G. J. Loake. *New Biotechnol.* **2014**, *31*, 242–245.

[2] S. Malik, R. M. Cusidó, M. H. Mirjalili, E. Moyano, J. Palazón, M. Bonfill. *Process Biochem.* **2011**, *46*, 23–34.

[3] H. Chen, X. Meng, X. Xu, W. Liu, S. Li. *Appl. Microbiol. Biotechnol.* **2018**, *102 (8)*, 3487–3495.

[4] A. Ghanem, H. Y. Aboul-Enein. *Tetrahedron Asymmetry* **2004**, *15 (21)*, 3331–3351.

[5] A. Ghanem. *Tetrahedron* **2007**, *63 (8)*, 1721–1754.

[6] K. d. G. Daiha, R. Angeli, S. D. d. Oliveira, R. V. Almeida. *PLOS ONE* **2015**, *10 (6)*, e0131624.

[7] B. Schulze, M. G. Wubbolts. *Curr. Opin. Biotechnol.* **1999**, *10 (6)*, 609–615.

[8] L. Casas-Godoy, F. Gasteazoro, S. Duquesne, F. Bordes, A. Marty, G. Sandoval, *Lipases: An Overview. In Lipases and Phospholipases* in *Springer Humana Press: New York, NY*, pp. 3–38.

[9] C. M. Clouthier, J. N. Pelletier. *Chem. Soc. Rev.* **2012**, *41 (4)*, 1585–1605.

[10] S. Soliman, Y. Tang. *Biotechnology and bioengineering* **2015**, *112*, 229–235.

[11] R. N. Patel. *Annu. Rev. Microbiol.* **1998**, *52*, 361–395.

[12] G. M. Cragg. *Med. Res. Rev.* **1998**, *18*, 315–331.

[13] R. Rabkin, D. P. Stables, N. W. Levin, M. M. Suzman. *The American journal of cardiology* **1966**, *18*, 370–380.

[14] E. Besterman, D. Friedlander. *Postgraduate medical journal* **1965**, *41*, 526.

[15] R. Čižmáriková, L. Habala, J. Valentová, M. Markuliak. *Applied Sciences* **2019**, *9*, 625.

[16] *Bristol-Myers Squibb trademarks*, https://www.bms.com/trademarks.html, Accessed: 2019-04-10.

[17] *Taxol injection label - Food and Drug Administration*, https://www.accessdata.fda.gov/drugsatfda_docs/-label/2011/020262s049lbl.pdf, Accessed: 2019-04-10.

[18] B. Zhang, A. Maiti, S. Shively, F. Lakhani, G. McDonald-Jones, J. Bruce, E. B. Lee, S. X. Xie, S. Joyce, C. Li, P. M. Toleikis, V. M.-Y. Lee, J. Q. Trojanowski. *Proc. Nat. Acad. Sci.* **2005**, *102*, 227–231.

[19] A. Ehrlich, S. Booher, Y. Becerra, D. L. Borris, W. Figg, M. L. Turner, A. Blauvelt. *Journal of the American Academy of Dermatology* **2004**, *50*, 533 – 540.

[20] P. Schrepfer, A. Buettner, C. Goerner, M. Hertel, J. v. Rijn, F. Wallrapp, W. Eisenreich, V. Sieber, R. Kourist, T. Brück. *Proc. Nat. Acad. Sci.* **2016**, *113*, E958–967.

[21] D. Stull, N. Jans, *Current taxol production from yew bark and future production strategies* in *2nd NCI Workshop on Taxol and Taxus, Arlington, VA*.

Bibliography

[22] N. Vidensek, P. Lim, A. Campbell, C. Carlson. *Journal of Natural Products* **1990**, *53*, 1609–1610.

[23] E. M. Croom Jr. *Taxol: Science and Applications* **1995**, 37–70.

[24] R. A. Holton, C. Somoza, H. B. Kim, F. Liang, R. J. Biediger, P. D. Boatman, M. Shindo, C. C. Smith, S. Kim. *Journal of the American Chemical Society* **1994**, *116*, 1597–1598.

[25] R. A. Holton, H. B. Kim, C. Somoza, F. Liang, R. J. Biediger, P. D. Boatman, M. Shindo, C. C. Smith, S. Kim. *Journal of the American Chemical Society* **1994**, *116*, 1599–1600.

[26] K. Nicolaou, Z. Yang, J. Liu, H. Ueno, P. Nantermet, R. Guy, C. Claiborne, J. Renaud, E. Couladouros, K. Paulvannan, et al.. *Nature* **1994**, *367*, 630.

[27] J. J. Masters, J. Link, L. B. Snyder, W. B. Young, S. J. Danishefsky. *Angewandte Chemie International Edition in English* **1995**, *34*, 1723–1726.

[28] P. A. Wender, N. F. Badham, S. P. Conway, P. E. Floreancig, T. E. Glass, J. B. Houze, N. E. Krauss, D. Lee, D. G. Marquess, P. L. McGrane, et al.. *Journal of the American Chemical society* **1997**, *119*, 2757–2758.

[29] I. Shiina, K. Saitoh, I. Fréchard-Ortuno, T. Mukaiyama. *Chemistry letters* **1998**, *27*, 3–4.

[30] K. Morihira, R. Hara, S. Kawahara, T. Nishimori, N. Nakamura, H. Kusama, I. Kuwajima. *Journal of the American Chemical society* **1998**, *120*, 12980–12981.

[31] T. Doi, S. Fuse, S. Miyamoto, K. Nakai, D. Sasuga, T. Takahashi. *Chemistry–An Asian Journal* **2006**, *1*, 370–383.

[32] A. Mendoza, Y. Ishihara, P. S. Baran. *Nature chemistry* **2012**, *4*, 21.

[33] D. Guenard, F. Gueritte-Voegelein, P. Potier. *Accounts of chemical research* **1993**, *26*, 160–167.

[34] E. Baloglu, D. G. Kingston. *Journal of natural products* **1999**, *62*, 1068–1071.

[35] M. E. Kolewe, V. Gaurav, S. C. Roberts. *Molecular pharmaceutics* **2008**, *5*, 243–256.

[36] S. Jennewein, R. Croteau. *Applied microbiology and biotechnology* **2001**, *57*, 13–19.

[37] D. Frense. *Applied microbiology and biotechnology* **2007**, *73*, 1233–1240.

[38] Y. J. Hong, D. J. Tantillo. *J. Am. Chem. Soc.* **2011**, *133*, 18249–18256.

[39] S. C. Roberts. *Nat. Chem. Biol.* **2007**, *3*, 387–395.

[40] A. E. Koepp, M. Hezari, J. Zajicek, B. S. Vogel, R. E. LaFever, N. G. Lewis, R. Croteau. *Journal of Biological Chemistry* **1995**, *270*, 8686–8690.

[41] M. Hezari, N. G. Lewis, R. Croteau. *Archives of Biochemistry and Biophysics* **1995**, *322*, 437–444.

[42] X. Lin, M. Hezari, A. E. Koepp, H. G. Floss, R. Croteau. *Biochemistry* **1996**, *35*, 2968–2977.

[43] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson. *Nature.* **2011**, *469*, 116–120.

[44] D. J. Tantillo. *Nat. Prod. Rep.* **2011**, *28*, 1035–1053.

[45] Y. Jin, D. C. Williams, R. Croteau, R. M. Coates. *J. Am. Chem. Soc.* **2005**, *127*, 7834–7842.

[46] Q. Jin, D. C. Williams, M. Hezari, R. Croteau, R. M. Coates. *J. Org. Chem.* **2005**, *70*, 4667–4675.

[47] S. Y. Chow, H. J. Williams, Q. Huang, S. Nanda, A. I. Scott. *J. Org. Chem.* **2005**, *70*, 9997–10003.

[48] D. C. Williams, B. J. Carroll, Q. Jin, C. D. Rithner, S. R. Lenger, H. G. Floss, R. M. Coates, R. M. Williams, R. Croteau. *Chem. Biol.* **2000**, *7*, 969–977.

[49] P. Gutta, D. J. Tantillo. *Org. Lett.* **2007**, *9*, 1069–1071.

[50] S. Edgar, F.-S. Li, K. Qiao, J.-K. Weng, G. Stephanopoulos. *ACS synthetic biology* **2016**, *6*, 201–205.

[51] T. A. Pemberton, M. Chen, G. G. Harris, W. K. W. Chou, L. Duan, M. Köksal, A. S. Genshaft, D. E. Cane, D. W. Christianson. *Biochemistry.* **2017**, *56*, 2010–2023.

[52] M. J. Begley, C. B. Jackson, G. Pattenden. *Tetrahedron* **1990**, *46*, 4907–4924.

[53] T. Tokiwano, T. Endo, T. Tsukagoshi, H. Goto, E. Fukushi, H. Oikawa. *Organic & biomolecular chemistry* **2005**, *3*, 2713–2722.

[54] D. C. Williams, M. R. Wildung, A. Q. Jin, D. Dalal, J. S. Oliver, R. M. Coates, R. Croteau. *Arch. Biochem. Biophys.* **2000**, *379*, 137–146.

[55] M. R. Wildung, R. Croteau. *Journal of Biological Chemistry* **1996**, *271*, 9201–9204.

[56] D. W. Christianson. *Chemical reviews* **2006**, *106*, 3412–3442.

[57] K. U. Wendt, G. E. Schulz. *Structure* **1998**, *6*, 127–133.

[58] D. W. Christianson. *Curr. Opin. Chem. Biol.* **2008**, *12*, 141–150.

[59] K. U. Wendt, G. E. Schulz, E. J. Corey, D. R. Liu. *Angewandte Chemie International Edition* **2000**, *39*, 2812–2833.

[60] M. J. Rynkiewicz, D. E. Cane, D. W. Christianson. *Proceedings of the National Academy of Sciences* **2001**, *98*, 13543–13548.

[61] J. Bohlmann, G. Meyer-Gauen, R. Croteau. *Proceedings of the National Academy of Sciences* **1998**, *95*, 4126–4133.

[62] D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau, D. W. Christianson. *Proc. Natl. Acad. Sci.* **2002**, *99*, 15375–15380.

[63] D. T. Major, M. Weitman. *J. Am. Chem. Soc.* **2012**, *134*, 19454–19462.

[64] M. W. v. d. Kamp, J. Sirirak, J. Żurek, R. K. Allemann, A. J. Mulholland. *Biochemistry.* **2013**, *52*, 8094–8105.

[65] C. M. Starks, K. Back, J. Chappell, J. P. Noel. *Science* **1997**, *277*, 1815–1820.

[66] A. M. Escorcia, J. P. M. v. Rijn, G. J. Cheng, P. Schrepfer, T. B. Brück, W. Thiel. *J. Comput. Chem.* **2018**, *39*, 1215–1225.

[67] M. Xu, P. R. Wilderman, R. J. Peters. *Proc. Natl. Acad. Sci.* **2007**, *104*, 7397–7401.

[68] H. Zhou, J. Chen, L. Ye, H. Lin, Y. Yuan. *Bioresour. Technol.* **2011**, *102 (10)*, 5562–5566.

[69] E. Y. Shishova, L. Di Costanzo, D. E. Cane, D. W. Christianson. *Biochemistry* **2007**, *46*, 1941–1951.

[70] D. C. Hyatt, B. Youn, Y. Zhao, B. Santhamma, R. M. Coates, R. B. Croteau, C. Kang. *Proc. Natl. Acad. Sci.* **2007**, *104*, 5360–5365.

Bibliography

[71] Y. J. Hong, D. J. Tantillo. *Org. Biomol. Chem.* **2010**, *8*, 4589–4600.

[72] Y. Freud, T. Ansbacher, D. T. Major. *ACS Catal.* **2017**, *7*, 7653–7657.

[73] T. Ansbacher, Y. Freud, D. T. Major. *Biochemistry.* **2018**, *57*, 3773–3779.

[74] D. Major, Y. Freud, M. Weitman. *Curr. Opin. Chem. Biol* **2014**, *21*, 25–33.

[75] M. Weitman, D. T. Major. *Journal of the American Chemical Society* **2010**, *132*, 6349–6360.

[76] M. Dixit, M. Weitman, J. Gao, D. T. Major. *ACS Catal.* **2017**, *7*, 812–818.

[77] D. T. Major. *ACS Catal.* **2017**, *7*, 5461–5465.

[78] L. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini. *Biochemistry* **1994**, *33*, 10871–10877.

[79] K. Stoschitzky, W. Lindner, G. Egginger, F. Brunner, B. Obermayer-Pietsch, A. Passath, W. Klein. *Clinical Pharmacology & Therapeutics* **1992**, *51*, 445–453.

[80] D. Patakas, V. Argiropoulou, G. Louridas, V. Tsara. *Thorax* **1983**, *38*, 108–112.

[81] W. Bode, A. E. Toet, A. A. Stolker, L. A. Van Ginkel, K. Groen, J. Wemer, D. J. De Wildt. *Chirality* **1995**, *7*, 626–631.

[82] A. E. Toet, A. V. De Kuil, W. Vleeming, J. Wemer, W. Bode, J. Meulenbelt, D. J. De Wildt. *Chirality* **1996**, *8*, 411–417.

[83] H. Bevinakatti, A. Banerji. *The Journal of Organic Chemistry* **1991**, *56*, 5372–5375.

[84] S. V. Darnle, P. N. Patil, M. M. Salunkhe. *Synthetic communications* **1999**, *29*, 3855–3862.

[85] A. Kamal, M. Sandbhor, A. Ali Shaik. *Bioorg. Med. Chem. Lett.* **2004**, *14 (17)*, 4581–4583.

[86] R. A. Veloo, G.-J. Koomen. *Tetrahedron Asymmetry* **1993**, *4*.

[87] H. Sasai, N. Itoh, T. Suzuki, M. Shibasaki. *Tetrahedron Lett.* **1993**, *34 (5)*, 855–858.

[88] Y.-F. Wang, S.-T. Chen, K. K.-C. Liu, C.-H. Wong. *Tetrahedron Lett.* **1989**, *30 (15)*, 1917–1920.

[89] O. Barbosa, C. Ariza, C. Ortiz, R. Torres. *New Biotechnol.* **2010**, *27 (6)*, 844–850.

[90] T.-W. Chiou, C.-C. Chang, C.-T. Lai, D.-F. Tai. *Bioorg. Med. Chem. Lett.* **1997**, *7 (4)*, 433–436.

[91] A. M. Escorcia, D. Molina, M. C. Daza, M. Doerr. *J. Mol. Catal. B Enzym.* **2013**, *98*, 21–29.

[92] Y.-Q. Wen, R. Hertzberg, I. Gonzalez, C. Moberg. *Chem. – Eur. J.* **2014**, *20 (13)*, 3806–3812.

[93] S. P. Panchgalle, R. G. Gore, S. P. Chavan, U. R. Kalkote. *Tetrahedron Asymmetry* **2009**, *20 (15)*, 1767–1770.

[94] A. Sikora, A. Tarczykowska, J. Chałupka, M. P. Marszałł. *Med. Res. J.* **2018**, *3 (1)*, 38–42.

[95] M. Romero, J. Gomez, B. Diaz-Suelto, A. Garcia-Sanz, N. Baster. *Appl. Biochem. Biotechnol.* **2011**, *165 (5)*, 1129–1140.

[96] M. Rodríguez-Mata, V. Gotor-Fernández, J. González-Sabín, F. Rebolledo, V. Gotor. *Org. Biomol. Chem.* **2011**, *9 (7)*, 2274–2278.

[97] D. Brady, S. Reddy, B. Mboniswa, L. H. Steenkamp, A. L. Rousseau, C. J. Parkinson, J. Chaplin, R. K. Mitra,

T. Moutlana, S. F. Marais, et al. *J. Mol. Catal. B Enzym.* **2012**, *75*, 1–10.

[98] B. Gao, T. Xu, J. Lin, L. Zhang, E. Su, Z. Jiang, D. Wei. *J. Mol. Catal. B Enzym.* **2011**, *68 (3–4)*, 286–291.

[99] Y. Zhang, J. Li, D. Han, H. Zhang, P. Liu, C. Li. *Biochem. Biophys. Res. Commun.* **2008**, *365 (4)*, 609–613.

[100] M. Ahmed, T. Kelly, A. Ghanem. *Tetrahedron* **2012**, *68 (34)*, 6781–6802.

[101] Z. S. Seddigi, M. S. Malik, S. A. Ahmed, A. O. Babalghith, A. Kamal. *Coord. Chem. Rev.* **2017**, *348*, 54–70.

[102] L. Piovan, M. D. Pasquini, L. H. Andrade. *Mol. Basel Switz.* **2011**, *16 (9)*, 8098–8109.

[103] C. Pilissão, P. d. O. Carvalho, M. d. G. Nascimento. *Process Biochem.* **2009**, *44 (12)*, 1352–1357.

[104] F. van Rantwijk, R. A. Sheldon. *Tetrahedron* **2004**, *60 (3)*, 501–519.

[105] Q. Wu, P. Soni, M. T. Reetz. *J. Am. Chem. Soc.* **2013**, *135 (5)*, 1872–1881.

[106] Z. Houiene, M. Merabet-Khelassi, N. Bouzemi, O. Riant, L. Aribi-Zouioueche. *Tetrahedron Asymmetry* **2013**, *24 (5)*, 290–296.

[107] M. Merabet-Khelassi, Z. Houiene, L. Aribi-Zouioueche, O. Riant. *Tetrahedron Asymmetry* **2012**, *23 (11)*, 828–833.

[108] H. Strohalm, S. Dold, K. Pendzialek, M. Weiher, K.-H. Engel. *J. Agric. Food Chem.* **2010**, *58 (10)*, 6328–6333.

[109] A. Alalla, M. Merabet-Khelassi, O. Riant, L. Aribi-Zouioueche. *Tetrahedron Asymmetry* **2016**, *27 (24)*, 1253–1259.

[110] H.-J. Wen, Q. Chen, G.-J. Zheng. *Chin. Chem. Lett.* **2015**, *26 (11)*, 1431–1434.

[111] T. Nielsen, M. Ishii, O. Kirk, *Lipases A and B from the yeast Candida antarctica* in *Biotechnological applications of cold-adapted organisms*, Springer, **1999**, pp. 49–61.

[112] H. Heldt-Hansen, M. Ishii, S. Patkar, T. Hansen, P. Eigtved, *new immobilized positional nonspecific lipase for fat modification and ester synthesis* in *ACS Symposium series-American Chemical Society*.

[113] J. Uppenberg, M. T. Hansen, S. Patkar, T. A. Jones. *Structure* **1994**, *2*, 293–308.

[114] J. Uppenberg, N. Oehrner, M. Norin, K. Hult, G. J. Kleywegt, S. Patkar, V. Waagen, T. Anthonsen, T. A. Jones. *Biochemistry* **1995**, *34*, 16838–16851.

[115] L. Hedstrom. *Chemical reviews* **2002**, *102*, 4501–4524.

[116] A. Rauwerdink, R. J. Kazlauskas. *Acs Catalysis* **2015**, *5*, 6153–6176.

[117] A. M. Escorcia, M. C. Daza, M. Doerr. *Journal of Molecular Catalysis B: Enzymatic* **2014**, *108*, 21–31.

[118] A. M. Escorcia, K. Sen, M. C. Daza, M. Doerr, W. Thiel. *ACS Catalysis* **2017**, *7*, 115–127.

[119] D. I. Barrera Valderrama, M. Doerr, M. C. Daza Espinosa, D. I. Barrera Valderrama, M. Doerr, M. C. Daza Espinosa. *Rev. Colomb. Biotecnol.* **2018**, *20 (1)*, 16–30.

[120] J. Ottosson, K. Hult. *J. Mol. Catal. B Enzym.* **2001**, *11 (4–6)*, 1025–1028.

[121] R. Chênevert, N. Pelchat, P. Morin. *Tetrahedron Asymmetry* **2009**, *20 (10)*, 1191–1196.

[122] L. Zhu, F. Zhu, S. Qin, B. Wu, B. He. *J. Mol. Catal. B Enzym.* **2016**, *133*, S150–S156.

Bibliography

[123] N. Melais, L. Aribi-Zouioueche, O. Riant. *Comptes Rendus Chim.* **2016**, *19 (8)*, 971–977.

[124] J. P. M. v. Rijn, A. M. Escorcia, W. Thiel. *Journal of computational chemistry* **2019**, *40*, 1902–1910.

[125] J. Nyhlén, B. Martín-Matute, A. G. Sandström, M. Bocola, J. Bäckvall. *ChemBioChem* **2008**, *9 (12)*, 1968–1974.

[126] T. Ema, Y. Nakano, D. Yoshida, S. Kamata, T. Sakai. *Org. Biomol. Chem.* **2012**, *10 (31)*, 6299–6308.

[127] S. Jiang, L. Zhang, D. Cui, Z. Yao, B. Gao, J. Lin, D. Wei. *Sci. Rep.* **2016**, *6*, 34750.

[128] H. J. Wijma, R. J. Floor, S. Bjelic, S. J. Marrink, D. Baker, D. B. Janssen. *Angew. Chem. Int. Ed.* **2015**, *54 (12)*, 3726–3730.

[129] R. Li, H. J. Wijma, L. Song, Y. Cui, M. Otzen, Y. Tian, J. Du, T. Li, D. Niu, Y. Chen, et al. *Nat. Chem. Biol.* **2018**, *14 (7)*, 664.

[130] A. Szabo, N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Courier Corporation, **2012**.

[131] D. J. Tantillo. *Angewandte Chemie International Edition* **2017**, *56*, 10040–10045.

[132] S. Tomić, M. Ramek. *Journal of Molecular Catalysis B: Enzymatic* **2006**, *38*, 139–147.

[133] Y. Zhang, J. Kua, J. A. McCammon. *The Journal of Physical Chemistry B* **2003**, *107*, 4459–4463.

[134] P.-O. Syrén, K. Hult. *ChemBioChem* **2010**, *11*, 802–810.

[135] G. G. Hammes, S. J. Benkovic, S. Hammes-Schiffer. *Biochemistry* **2011**, *50*, 10422–10430.

[136] K. Arora, C. L. Brooks, *Multiple intermediates, diverse conformations, and cooperative conformational changes underlie the catalytic hydride transfer reaction of dihydrofolate reductase* in *Dynamics in Enzyme Catalysis*, Springer, **2013**, pp. 165–187.

[137] M. Garcia-Viloca, J. Gao, M. Karplus, D. G. Truhlar. *Science* **2004**, *303*, 186–195.

[138] S. Martí, M. Roca, J. Andrés, V. Moliner, E. Silla, I. Tuñón, J. Bertran. *Chemical Society Reviews* **2004**, *33*, 98–107.

[139] J. Villá, A. Warshel. *The Journal of Physical Chemistry B* **2001**, *105*, 7887–7907.

[140] A. Warshel. *Annual Review of Biophysics and Biomolecular Structure* **2003**, *32*, 425–443.

[141] X. Zhang, K. N. Houk. *Accounts of Chemical Research* **2005**, *38*, 379–385.

[142] D. A. McQuarrie, *Quantum chemistry*, University Science Books, **2008**.

[143] D. J. Griffiths, *Introduction to quantum mechanics*, Pearson international edition (Pearson Prentice Hall, 2005), **2010**.

[144] P. W. Atkins, J. De Paula, *Atkins' physical chemistry*, Oxford university press, **2006**.

[145] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. W. rkiewicz Kuczera, D. Yin, M. Karplus. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[146] A. D. Mackerell, M. Feig, C. L. Brooks. *J. Chem.* **2004**, *25*, 1400–1415.

[147] M. Born, R. Oppenheimer. *Annalen der physik* **1927**, *389*, 457–484.

[148] C. Møller, M. S. Plesset. *Physical review* **1934**, *46*, 618.

[149] G. D. Purvis III, R. J. Bartlett. *The Journal of Chemical Physics* **1982**, *76*, 1910–1918.

[150] P. Hohenberg, W. Kohn. *Physical review B* **1964**, *136*, 864–871.

[151] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais. *Phys. Rev. B* **1992**, *46*, 6671–6687.

[152] D. C. Langreth, M. J. Mehl. *Phys. Rev. B* **1983**, *28*, 1809–1834.

[153] J. P. Perdew, K. Burke, M. Ernzerhof. *Physical review letters* **1996**, *77*, 3865.

[154] A. D. Becke. *Physical review A* **1988**, *38*, 3098.

[155] A. D. Becke. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[156] M. J. Dewar, W. Thiel. *Journal of the American Chemical Society* **1977**, *99*, 4899–4907.

[157] D. Bakowies, W. Thiel. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

[158] I. Antes, W. Thiel, *On the treatment of link atoms in hybrid methods* in *ACS Symposium Series, Vol. 712: Combined Quantum Mechanical and Molecular Mechanical Methods*, American Chemical Society., pp. 50–65.

[159] P. Sherwood, A. H. d. Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, e. al. *Mol. Struct.: Theochem.* **2003**, *632*, 1–28.

[160] P. P. Ewald. *Annalen der physik* **1921**, *369*, 253–287.

[161] D. Beglov, B. Roux. *The Journal of chemical physics* **1994**, *100*, 9050–9063.

[162] D. Frenkel, B. Smit, *Understanding molecular simulation: From algorithms to applications*, Computational sciences series 1, **2002**.

[163] A. Banerjee, N. Adams, J. Simons, R. Shepard. *The Journal of Physical Chemistry* **1985**, *89*, 52–57.

[164] J. Baker. *Journal of Computational Chemistry* **1986**, *7*, 385–395.

[165] L. Deng, T. Ziegler, L. Fan. *The Journal of chemical physics* **1993**, *99*, 3823–3835.

[166] L. Deng, T. Ziegler. *International Journal of Quantum Chemistry* **1994**, *52*, 731–765.

[167] K. M. Ishida, Kazuhiro, A. Komornicki. *The Journal of chemical physics* **1977**, *66*, 2153–2156.

[168] J. Decius, P. C. Cross, E. B. Wilson, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, Dover Publications, **1980**.

[169] R. W. Zwanzig. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

[170] G. M. Torrie, J. P. Valleau. *Journal of Computational Physics* **1977**, *23*, 187–199.

[171] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman. *Journal of computational chemistry* **1992**, *13*, 1011–1021.

[172] B. Christelle, B. D. O. Eduardo, C. Latifa, M. Elaine-Rose, M. Bernard, R.-H. Evelyne, G. Mohamed, E. Jean-Marc, H. Catherine. *Journal of biotechnology* **2011**, *156*, 203–210.

[173] P. Trodler, R. D. Schmid, J. Pleiss. *BMC structural biology* **2009**, *9*, 38.

[174] P. Trodler, J. Pleiss. *BMC Structural Biology* **2008**, *8*, 9.

[175] L. S. Caves, J. D. Evanseck, M. Karplus. *Protein Science* **1998**, *7*, 649–666.

[176] G. A. Worth, F. Nardi, R. C. Wade. *The Journal of Physical Chemistry B* **1998**, *102*, 6260–6272.

[177] B. Karasulu, M. Patil, W. Thiel. *J. Am. Chem. Soc.* **2013**, *135*, 13400–13413.

[178] E. Sanchez-Garcia, M. Doerr, Y.-W. Hsiao, W. Thiel. *The Journal of Physical Chemistry B* **2009**, *113*, 16622–16631.

[179] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

[180] P. H. Hünenberger, *Thermostat algorithms for molecular dynamics simulations* in *Advanced computer simulation*, Springer, **2005**, pp. 105–149.

[181] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert. *Physical Review B* **1998**, *58*, 7260.

[182] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus. *The Journal of Physical Chemistry B* **2001**, *105*, 569–585.

[183] R. Stote, M. Karplus, et al.. *Journal de chimie physique* **1991**, *88*, 2419–2433.

[184] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen. *Journal of computational physics* **1977**, *23*, 327–341.

[185] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker. *Nucleic acids research* **2004**, *32*, W665–W667.

[186] M. H. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen. *Journal of chemical theory and computation* **2011**, *7*, 525–537.

[187] C. R. Søndergaard, M. H. Olsson, M. Rostkowski, J. H. Jensen. *Journal of chemical theory and computation* **2011**, *7*, 2284–2295.

[188] P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll, J. S. Dickschat. *Angewandte Chemie International Edition* **2014**, *53*, 7652–7656.

[189] Z. Smedarchina, W. Siebrand, A. Fernández-Ramos, Q. Cui. *Journal of the American Chemical Society* **2003**, *125*, 243–251.

[190] B. K. Mai, K. Park, M. P. T. Duong, Y. Kim. *The Journal of Physical Chemistry B* **2012**, *117*, 307–315.

[191] D. Riccardi, P. König, X. Prat-Resina, H. Yu, M. Elstner, T. Frauenheim, Q. Cui. *Journal of the American Chemical Society* **2006**, *128*, 16302–16311.

[192] S. Metz, W. Thiel. *J. Am. Chem. Soc.* **2009**, *131*, 14885–14902.

[193] I. Polyak, M. T. Reetz, W. Thiel. *J. Am. Chem. Soc.* **2012**, *134*, 2732–2741.

[194] ChemShell, a Computational Chemistry Shell, see www.chemshell.org.

[195] Y. Zhao, D. G. Truhlar. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[196] A. Schäfer, C. Huber, R. Ahlrichs. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

[197] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, et al.. *Inc., Wallingford CT* **2013**.

[198] W. Smith, T. Forester. *Journal of molecular graphics* **1996**, *14*, 136–141.

[199] J. Nocedal. *Math. Comput.* **1980**, *35*, 773–782.

[200] D. C. Liu, J. Nocedal. *Math. Prog.* **1989**, *45*, 503–528.

[201] D. W. Christianson. *Chem. Rev.* **2017**, *117*, 11570–11648.

[202] M. Chen, W. K. Chou, N. Al-Lami, J. A. Faraldos, R. K. Allemann, D. E. Cane, D. W. Christianson. *Biochemistry* **2016**, *55*, 2864–2874.

[203] Y. Gao, R. B. Honzatko, R. J. Peters. *Natural product reports* **2012**, *29*, 1153–1175.

[204] D. E. Cane, H. T. Chiu, P. Liang, K. S. Anderson. *Biochemistry.* **1997**, *36*, 8332–8339.

[205] J. R. Mathis, K. Back, C. Starks, J. Noel, C. D. Poulter, J. Chappell. *Biochemistry.* **1997**, *36*, 8340–8348.

[206] M. Dixit, M. Weitman, J. Gao, D. T. Major. *ACS Catalysis* **2018**, *8*, 1371–1375.

[207] D. I. Barrera Valderrama, M. Doerr, M. C. Daza Espinosa. *Revista Colombiana de Biotecnología* **2018**, *20*, 16–30.

[208] O. Trott, A. J. Olson. *Journal of computational chemistry* **2010**, *31*, 455–461.

[209] M. W. Chang, C. Ayeni, S. Breuer, B. E. Torbett. *PloS one* **2010**, *5*, e11955.

[210] R. Ávila-González, M. Pérez-Gilabert, F. García-Carmona. *Journal of bioscience and bioengineering* **2005**, *100*, 423–428.

[211] S. J. Teague, A. M. Davis, P. D. Leeson, T. Oprea. *Angewandte Chemie International Edition* **1999**, *38*, 3743–3748.

[212] C. Oger, Z. Marton, Y. Brinkmann, V. Bultel-Poncé, T. Durand, M. Graber, J.-M. Galano. *The Journal of organic chemistry* **2010**, *75*, 1892–1897.

[213] E. B. De Oliveira, C. Humeau, L. Chebil, E. R. Maia, F. Dehez, B. Maigret, M. Ghoul, J.-M. Engasser. *Journal of Molecular Catalysis B: Enzymatic* **2009**, *59*, 96–105.

[214] Y. Cao, S. Wu, J. Li, B. Wu, B. He. *Journal of Molecular Catalysis B: Enzymatic* **2014**, *99*, 108–113.

[215] H. M. Senn, W. Thiel. *Angewandte Chemie International Edition* **2009**, *48*, 1198–1229.

[216] M. W. van der Kamp, A. J. Mulholland. *Biochemistry* **2013**, *52*, 2708–2728.

[217] T. E. O'Brien, S. J. Bertolani, D. J. Tantillo, J. B. Siegel. *Chemical science* **2016**, *7*, 4009–4015.

[218] T. E. O'Brien, S. J. Bertolani, Y. Zhang, J. B. Siegel, D. J. Tantillo. *ACS catalysis* **2018**, *8*, 3322–3330.

Bibliography

[219] B.-X. Tian, F. H. Wallrapp, G. L. Holiday, J.-Y. Chow, P. C. Babbitt, C. D. Poulter, M. P. Jacobson. *PLoS computational biology* **2014**, *10*.

[220] D. J. Tantillo. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, *10*, e1453.

[221] S. Das, M. Shimshi, K. Raz, N. Nitoker Eliaz, A. R. Mhashal, T. Ansbacher, D. T. Major. *Journal of chemical theory and computation* **2019**, *15*, 5116–5134.

A. Molecular Dynamics Study of Taxadiene Synthase Catalysis

Andrés M. Escorcia, **Jeaphianne P. M. van Rijn**, Gui-Juan Cheng, Patrick Schrepfer, Thomas B. Bruck and Walter Thiel

I carried out the MD calculations and performed the subsequent analysis for two of the four cations. I performed the analysis of the bond distances of all cations. I contributed to the writing of the manuscript.

Molecular Dynamics Study of Taxadiene Synthase Catalysis

Andrés M. Escorcia,[a] Jeaphianne P. M. van Rijn,[a] Gui-Juan Cheng,[a] Patrick Schrepfer,[b] Thomas B. Brück,[b] and Walter Thiel ●*[a]

Molecular dynamics (MD) simulations have been performed to study the dynamic behavior of noncovalent enzyme carbocation complexes involved in the cyclization of geranylgeranyl diphosphate to taxadiene catalyzed by taxadiene synthase (TXS). Taxadiene and the observed four side products originate from the deprotonation of carbocation intermediates. The MD simulations of the TXS carbocation complexes provide insights into potential deprotonation mechanisms of such carbocations. The MD results do not support a previous hypothesis that carbocation tumbling is a key factor in the deprotonation of the carbocations by pyrophosphate. Instead water bridges are identified which may allow the formation of side products via multiple proton transfer reactions. A novel reaction path for taxadiene formation is proposed on the basis of the simulations. © 2018 Wiley Periodicals, Inc.

Introduction

The enzyme taxadiene synthase (TXS), a class I terpene synthase, catalyzes the cyclization of the universal diterpene precursor (*E,E,E*)-geranylgeranyl diphosphate (GGPP) to taxa-4(5),11(12)-diene (taxadiene, **T**; see Fig. 1).[1–5] This transformation is the first committed step in the biosynthesis of taxol, a natural tetracyclic diterpene (originally isolated from the pacific yew tree *Taxus brevifolia*), which has potent anticancer activity and is clinically applied in the first line treatment of breast, lung, and ovarian cancer.[6–9] Understanding the molecular basis of TXS catalysis constitutes an active field of research,[3,5,10] as it may enable sustainable biotechnological pathways for taxol production that could substitute conventional semi-synthetic routes, which are associated with significant costs and the accumulation of significant toxic waste streams.[8,9,11]

Full-length TXS contains 862 residues; including an N-terminal transit sequence of approximately 80 residues, which is cleaved on plastidial protein maturation.[2,12] The catalytic (mature) TXS assembles into three α-helical domains. The active site of the enzyme is located in the C-terminal domain (S553-V862) containing the metal-binding motifs D^{613}DMAD (at helix D) and N^{757}DTKTYQAE (at helix H), which coordinate three Mg^{2+} ions (referred to as Mg$_A^{2+}$, Mg$_B^{2+}$, and Mg$_C^{2+}$ ions). The Mg$_A^{2+}$ and Mg$_C^{2+}$ ions are linked to D613 and D617, while N757, T761, and E765 coordinate to the Mg$_B^{2+}$ ion (see Fig. 2).[5] As TXS is a class I terpene synthase, it initiates catalysis via a trinuclear cluster of Mg^{2+} ions, which bind and trigger ionization of the GGPP diphosphate group (PPi). The PPi hydrolysis process leads to formation of the primary, highly reactive geranylgeranyl carbocation, which is the starting point of a complex reaction series toward taxadiene. To that end, the carbocation undergoes multistep rearrangement and cyclization reactions on route to taxadiene isomers.[5,15]

Several experimental and computational studies have been reported which provide insights into TXS catalysis.[3,4,10,16–21]

However, the catalytic role of the enzyme along the entire reaction path (from GGPP to **T**) is not yet defined in detail.

Quantum mechanical (QM) model calculations in the gas phase on the hydrocarbon part of GGPP have been used to propose a reaction mechanism for the cyclization of GGPP to **T**, which involves the formation of seven carbocation intermediates and is hereafter referred to as the HT-QM mechanism (HT being a label for its authors, see Fig. 1).[10] The HT-QM mechanism is supported by previous mechanistic experiments carried out with GGPP analogues as well as labeled GGPP, indicating the existence of carbocations **B**, **C**, **D**, and **E** during TXS catalysis (see Fig. 1).[4,17–21] Moreover, like other class I terpene synthases, TXS has been shown to be a promiscuous enzyme and apart from the major product **T** it produces four minor products: verticillia-3(4),7(8),12(13)-triene (**V**), verticillia-4(20),7(8),11(12)-triene (**V1**), verticillia-3(4),7(8),11(12)-triene (**V2**), and taxa-4(20),11(12)-diene (**T1**).[2,3,18,22] The formation of these side products supports the HT-QM mechanism, as they correspond to the deprotonation products of the cations **C-F** (see Fig. 1).

In a recent study,[3] an integrated approach of docking and molecular mechanics (MM) optimization was used to model the catalytically relevant closed complexes of TXS with GGPP as well as the carbocation intermediates proposed in the HT-QM mechanism. For the first time the TXS-GGPP complex was modeled in the closed active-site conformation, as class I terpene synthases are known to undergo a conformational change from an open to a closed active-site conformation on

[a] A. M. Escorcia, J. P. M. van Rijn, G.-J. Cheng, W. Thiel
Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, Mülheim 45470, Germany
E-mail: thiel@kofo.mpg.de
[b] P. Schrepfer, T. B. Brück
Professorship of Industrial Biocatalysis, Department of Chemistry, Technical University Munich, Lichtenberg Str. 4, Garching 85748, Germany
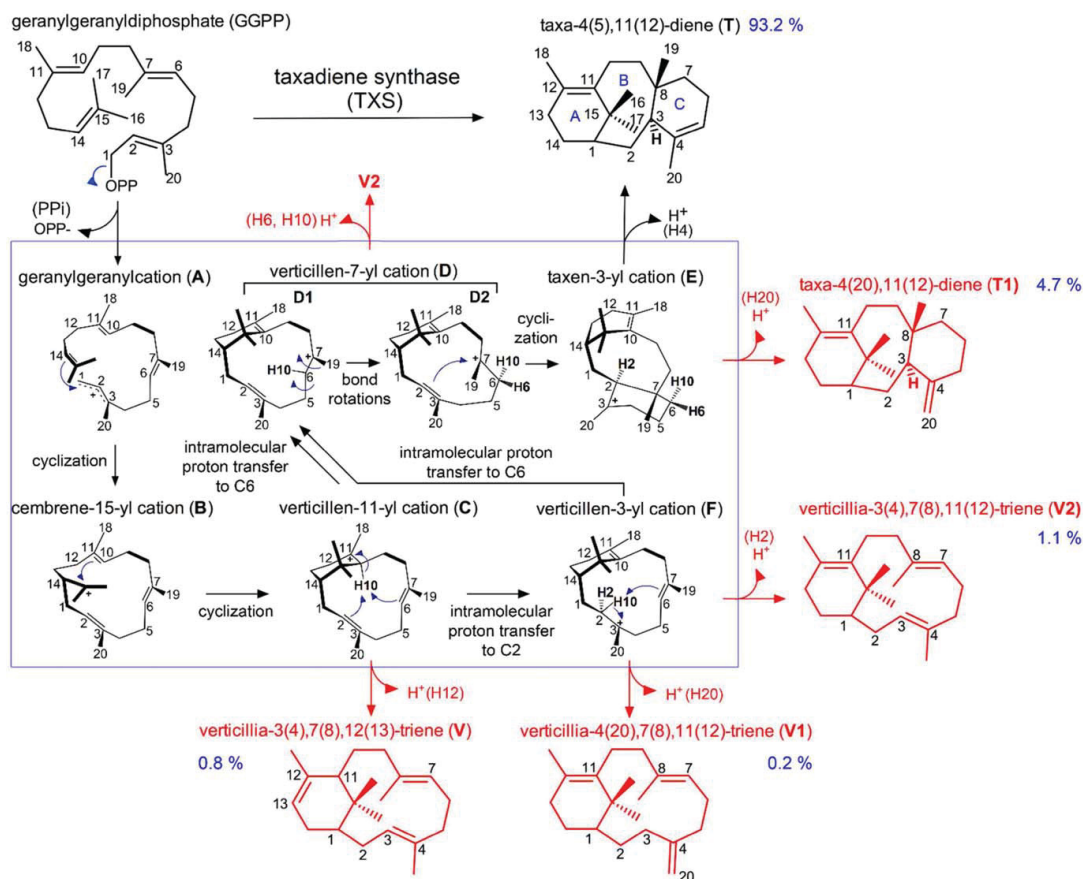Contract grant sponsor: Max Planck Society

Figure 1. In black: the reaction mechanism of the TXS-catalyzed cyclization of GGPP to taxadiene proposed on the basis of QM model calculations (HT-QM mechanism).[10] In red: minor products of TXS catalysis. Product yields (numbers in blue) are given as reported by Schrepfer et al.[3] The numbering of GGPP is used for the carbocation intermediates, while minor products are numbered according to taxadiene convention. OPP denotes the diphosphate group of GGPP. A–C labels in blue denote taxadiene ring systems. [Color figure can be viewed at wileyonlinelibrary.com]

binding the substrate and catalytic metal ions.[5,13,23–26] In this respect, closure of the active site protects highly reactive carbocation intermediates from premature quenching by bulk solvent.[15,25] In the available crystal structure of TXS (PDB ID 3P5R) the active site is not fully closed,[5] and hence homology modeling was applied to build a model of closed TXS,[3] using the crystal structure of the closed form of bornyl diphosphate synthase (BPPS, PDB ID 1N20)[13] as a template. In the resulting structure of closed TXS the active site is capped by three loops (namely, A–C (G570–H579), J–K (F837–E846), and H–α1 (Q770–S773)), the C-terminal portion of helix H (K760–G769), and the N terminus random coil segment (hereafter referred to as NTRC segment, D80–L108; see Fig. 2). Overall, the results obtained show that the HT-QM mechanism is feasible in the TXS environment, as the carbocation structures fit in the active site of TXS in a proper orientation such that each **TXS·cation** complex is distinctly derived from its precursor in the reaction

cascade (see Fig. 1).[3] Like in a previous report on other class I terpene synthases,[27] the complexes identified suggest that the pyrophosphate anion (PPi) plays a key role for TXS catalysis[3]; PPi is released by initial GGPP ionization and considered to be retained in the active site during the entire cyclization cascade.[13,23,27–29] The electrostatic interactions between PPi and the carbocations are expected to have a large influence on the thermodynamics of the reaction.[3] In addition, PPi is proposed to deprotonate the carbocations **C**, **F**, and **E**, thus being responsible for the formation of the side products.[3]

Besides PPi, an arginine residue (R580) located at the active site of TXS has also been suggested to act as a base and to deprotonate cation **C**.[3] However, the interatomic distances observed between some of the hydrogen atoms expected to be abstracted from the carbocations and either R580 or PPi are very large (3.7–4.8 Å), and carbocation tumbling has been

73

WWW.C-CHEM.ORG   FULL PAPER
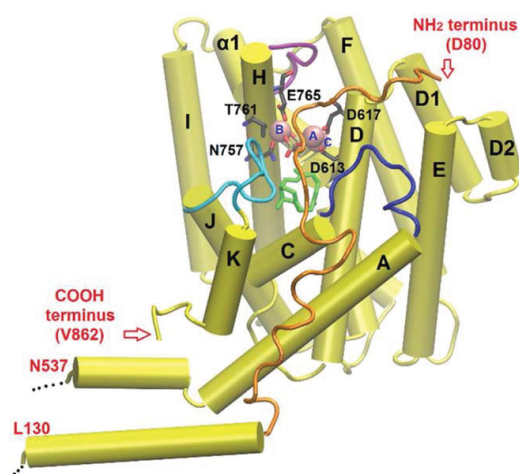
Figure 2. Representative structure of TXS in the closed active site conformation. Atomic coordinates were provided by Schrepfer et al.[3] The protein regions capping the active site are shown in different colors: loops A–C in blue, J–K in cyan, and H–α1 in purple. The N terminus random coil (NTRC) segment is shown in orange, while the C-terminal portion of helix H is shown in yellow like the rest of the protein. In licorice: residues acting as metal ligands (carbons in black), cation **C** (green), and PPi (phosphorus in tan). The catalytic Mg_{A-C}^{2+} ions are shown as spheres with their respective labels in blue. Hydrogen atoms are omitted for clarity. The labels for helical segments are based on farnesyl diphosphate synthase convention.[13,14] [Color figure can be viewed at wileyonlinelibrary.com]

then assumed to be involved in the deprotonation.[3] In general, all complexes encountered after **TXS·C** formation are proposed to undergo carbocation tumbling, which is thought to be possible due to three factors: (i) the oversized active-site cavity of TXS,[5] (ii) the electrostatic attraction between the cations and PPi,[3] and (iii) the smaller stabilization of cations **C–F** by the surrounding protein residues compared with cations **A–B** (based on the type of interactions involved).[3]

In the present study, we aim at gaining a more detailed understanding of TXS catalysis, by addressing the dynamic behavior of the complexes **TXS·C** to **TXS·F**. Starting from the structures obtained by Schrepfer et al.,[3] we carry out multiple 1 ns MD simulations in explicit water using different initial velocity distributions.[30–35] Through an extensive sampling of the system, these MD simulations allow us to assess the hypothesis of carbocation tumbling and to explore the probabilities for accessing different reaction pathways leading to the formation of taxadiene and reaction side products. Enzyme reactions are known to involve multiple configurations of enzyme-substrate complexes and an extensive sampling of the configurational space is required to fully understand the origin of the catalytic properties (activity, specificity, and selectivity) of enzymes.[33,36–39] Thus, our MD simulations aim at providing new insights into the molecular basis of TXS promiscuity and of the preference for taxadiene formation. In addition, they may serve as starting point for subsequent quantum mechanics/molecular mechanics (QM/MM) studies on the mechanism of TXS catalysis.[33,40–42]

Methods and Computational Details

The structures of the complexes between TXS and the cations **C** to **F** obtained by Schrepfer et al.[3] were used as starting structures in our MD study. Regarding cation **D**, we focused on **D1** (see Fig. 1). The **TXS·cation** complexes consisted of the closed TXS, the corresponding carbocation, the catalytic Mg^{2+} ions, and PPi. TXS was represented by the structural motifs directly involved in the catalysis, that is, the C-terminal domain (S553–V862) and part of the N-terminal domain (D80–L130 and N537–Q552; see Fig. S1a–b in the Supporting Information). In these structures, all acidic TXS residues were deprotonated (negatively charged), while lysine and arginine residues were protonated (positively charged). All histidine residues were singly protonated at N_δ except H740, which was doubly protonated.

These complexes were solvated in a water sphere with 37 Å radius cut out from an equilibrated simulation of TIP3P water under periodic boundary conditions. The sphere was centered at the $C_\gamma 1$ atom carbon of the residue V610 and covered almost the entire system (see Fig. S1c in the Supporting Information). All added water molecules whose oxygen atoms were within 2.8 Å of any existing nonhydrogen atom were eliminated. Using the Autoionize plugin of the VMD program[43] 10 Na^+ ions were randomly added to the water sphere (distant from the protein) to neutralize the system. The water molecules and Na^+ ions (together hereafter referred to as solvation layer) were geometry-optimized, by first performing 1000 steps of steepest descent (SD) followed by 1000 steps of adapted-basis Newton–Raphson minimization (ABNR), keeping all other atoms fixed. Thereafter, an active region was defined including all residues within 27 Å of the $V610:C_\gamma 1$ atom, PPi, the Mg^{2+} ions, and the carbocation. All protein atoms outside the active region were kept fixed in all subsequent calculations (1647 of 6133 protein atoms, see section 1 in the Supporting Information for details). The active region and the solvation layer were optimized, applying a harmonic positional restraint with force constant $k = 30$ kcal mol^{-1} $Å^{-2}$ to the former (two successive SD and ABNR minimizations, 1000 steps each). This was followed by a heating MD simulation (15,000 steps) starting at 50 K and ending at 300 K, raising the temperature in steps of 10 K every 100 MD steps; the Verlet algorithm was used with a time step of 1 fs. This solvation procedure was repeated 12 times, and in the last two cycles the number of steps used in the dynamics was increased to 30,000. The harmonic positional restraint applied to the active region was successively lowered in each of the solvation cycles. Finally, all the constraints were removed and a productive MD simulation of 1 ns at 300 K was carried out. A Langevin thermostat with a 1 ps^{-1} damping coefficient was employed to maintain a constant temperature.[44] The simulation system consisted of 23,641 atoms (TXS: 6133; cation: 53; PPi: 9; Mg^{2+}: 3; Na^+: 10; water: 17,433). The chosen MD setup is analogous to that of previous studies reported in the literature.[32,33,41,45]

Three MD simulations with different initial velocity distributions (referred to as MD1, MD2, and MD3) were performed for each complex to enhance sampling.[30–35] All MD simulations were done with the CHARMM software package (version 35b2).[46] We applied a QM/MM MD approach.[23,32,33] The QM

region (i.e., the carbocation) was treated by the self-consistent-charge density functional tight binding (SCC-DFTB) method.[47] The MM region (comprised of TXS, PPi, catalytic Mg^{2+} ions, and the solvation layer) was described by the CHARMM27 force field.[48,49] The QM/MM electrostatic interactions were calculated within the QM Hamiltonian as the Coulomb interaction between the Mulliken charges of the QM atoms and the MM point charges, as implemented for the SCC-DFTB method in the CHARMM program.[50] No cutoffs were applied for the nonbonded QM/MM interactions. The group-based extended electrostatics approach was used within the MM region;[51] the electrostatic interactions between particles closer than 14 Å were treated by the conventional pairwise additive scheme, while the interactions at larger distance were approximated by a computationally cheaper multipole approach. A quartic spherical boundary potential was applied to the water molecules to keep the shape of the water sphere and to prevent evaporation of outer water molecules.[46] All distances involving hydrogen atoms were constrained by SHAKE.[52]

The time evolution of the temperature and the potential energy were analyzed to assess the equilibration of the 1 ns MD trajectories. In all simulations, the system was equilibrated after 200 ps, so the remaining 800 ps were used for analysis. This analysis focused on the following: (i) the stability of the protein structure in closed conformation, (ii) the conformational and positional stability of the carbocations, and (iii) the potential reactivity of the carbocations. As a measure of the structural stability of the protein, we analyzed the time evolution of the root mean square deviation (RMSD; computed with respect to the initial structure) for all heavy atoms of the protein backbone. In addition, we calculated the root mean square fluctuation (RMSF) per residue to identify the regions of the protein that display high and reduced flexibility. RMSD calculations with respect to the initial structure were also carried out for all carbon atoms of the carbocations to investigate their conformational and positional stability at the active site of TXS. The structures of the carbocations during the MD simulations were subjected to cluster analysis using the GROMOS clustering algorithm.[53] The potential reactivity of the carbocations was qualitatively analyzed by considering the average values of the catalytically relevant interatomic distances (e.g., the distance between the atoms C2 and H10 of cation **C**, see Fig. 1).[32,54] Furthermore, to assess the chances of the side chains of active-site residues to be involved in the deprotonation of the carbocations, we used the PROPKA module of the PDB2PQR server to predict their pKa values.[55–57] The latter were computed for the protein alone; potential shifts of the pKa values due to the presence of PPi, the carbocation, and the Mg^{2+} ions were qualitatively assessed by visual inspection of the surroundings of these residues in the **TXS·cation** complexes.

Results and Discussion

Structural stability of the protein

The structure of closed TXS is well conserved throughout all MD simulations. The average RMSD values of all heavy atoms of the protein backbone with respect to the initial structure are between 0.97 and 1.2 Å, with standard deviations of $\leq$0.2 Å (see Table S1 in the Supporting Information). The catalytic Mg^{2+} ions stay simultaneously coordinated to PPi and the respective ligating residues of helices H and D, and thus, the position of PPi near the active site is stable along the MD trajectories (see Fig. S2 and Table S2 in the Supporting Information). Protein regions with higher flexibility include the J–K and A–C loops and the NTRC segment, the latter being most flexible with residues exhibiting RMSF values of >1.6 Å (see Fig. 3 and Figs. S2–S3 in the Supporting Information). Nevertheless, these regions stay in a proper orientation to enclose the active site during the MD simulations. The residues Y89, H90, D92, Y841, and N845 play a key role in this respect. The polar side chains of H90, D92, and N845 mediate the formation of hydrogen bonds between the flexible regions (involving the pairs of residues D92–H579, A844–H90, and N845–H90) keeping them close to each other during the MD simulations. Meanwhile, throughout all trajectories, Y89 forms hydrogen bonds with R580 (helix C) and/or D614 (helix D), and/or a water-mediated hydrogen bond with PPi. In addition, for more than 94% of the simulation time Y841 forms π-stacking interactions with F834 (helix J) and a hydrogen bond with S587 (helix C). Furthermore, a water-mediated hydrogen bond between Y841 and R580 is present for more than 53% of the simulation time in at least one of the three MD simulations of each **TXS·cation** complex. The interactions of Y89 and Y841 pull the NTRC segment and the J–K loop toward the catalytic cavity of TXS (see Fig. 4 and Table S3 in the Supporting Information). In contrast, the C-terminal portion of helix H and the H-α1 loop are less flexible (see Fig. 3 and Fig. S2 in the Supporting Information). These polypeptide segments stay oriented toward the catalytic cavity during the MD simulations thanks to (i) the coordination of the residues T761 and E765 to the $D_{614,617}$-$Mg^{2+}_{A,B}$-PPi-Mg^{2+}_{B} motif and (ii) the formation of water-mediated hydrogen bonds between PPi and the residues R768 and Q770 (see Fig. 4 and Table S3 in the Supporting Information). Thus, the active site of TXS is shielded from the bulk solvent throughout all MD simulations, which excludes solvent participation in the deprotonation of the carbocations.

Conversely, the active-site cavity harboring the carbocations is rigid, as indicated by RMSF values of 0.3–0.8 Å obtained for most of its residues (see Fig. 3). This restricts the positional freedom of the carbocations, and hence the promiscuity of TXS should not be attributed to carbocation tumbling (see below). Overall, two factors appear to be the main contributors to the rigidity of this cavity: (i) the ability of its polar residues to retain very strong (multiple) interactions with their surroundings during the MD simulations (e.g., Y835 forms simultaneously a water-mediated hydrogen bond with PPi and a hydrogen bond with N757 during 68–92% of the simulation time), and (ii) the orientation of residues with bulky side chains toward the interior of the catalytic cavity such that there is not enough space for them to move freely. See Table S3 and Figure S4 of the Supporting Information for more details.
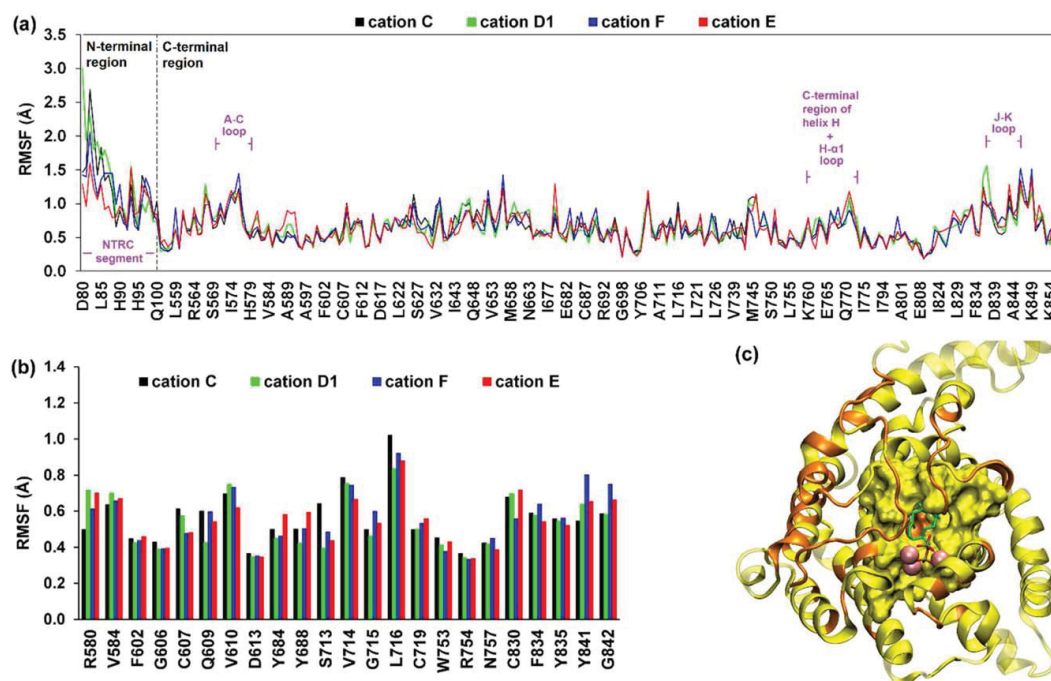
Figure 3. a) RMSF per residue values averaged over three MD simulations with different seed velocities. Residues kept fixed during the simulations are omitted (see section 1 of the Supporting Information for details). b) Expanded plot of the RMSF values for the residues comprising the cavity of TXS harboring the carbocations. c) Representative structure of the system: the regions of the protein displaying higher flexibility are highlighted in orange, while the cavity harboring the carbocations is shown in surface representation. [Color figure can be viewed at wileyonlinelibrary.com]

The water-mediated hydrogen bonds aforementioned are possible because of the presence of water molecules at the entrance of the active site, which is enriched in polar residues and harbors the catalytic Mg^{2+} ions and PPi. This is consistent with the reported crystal structures of class I terpene synthases in closed form, in which the Mg^{2+} ions and PPi are surrounded by a significant number of water molecules ($\sim$13–17) showing that some solvent molecules can remain inside the active site after its closure (see e.g., the PDB entries 2OA6, 1N23, 1N20, and 4OKZ).[13,58,59] The importance of PPi and Mg^{2+} ions for the closed conformation of such enzymes is supported not only by X-ray crystallography studies but also by recent MD work on aristolochene synthase.[15,24,25]

The results of the MD simulations on the dynamic behavior and potential reactivity of the carbocations in the active site of TXS are discussed in the following sections. There is one section for every **TXS·cation** complex.

TXS·C complex

Both the conformation and orientation of cation **C** inside the active site of TXS are very stable throughout the MD simulations of the **TXS·C** complex. The dominant conformation of cation **C** during the MD simulations is equivalent to the HT-QM structure (see Fig. 5 and Table S4 in the Supporting Information). This reinforces previous conclusions on the feasibility

of the HT-QM mechanism in the TXS environment.[3] The 6-membered ring (A ring) of cation **C** is in a chair-like conformation along all trajectories, and its 12-membered ring undergoes only slight fluctuations around the dominant structure. Hence, part of the catalytic role of TXS is to restrict the conformational freedom of cation **C,** such that there is a preference for the conformer suitable for rearranging to the subsequent carbocation intermediates on route to taxadiene. This is in full agreement with previous experiments carried out on the chemical transformation of verticillene and its derivatives, in which the nonenzymatic transformation of cation **C** to taxadiene was unsuccessful or proceeded with a very low yield (0.004%).[60,61] The average value of the relevant interatomic distance for the intramolecular proton transfer leading to formation of cation **F** (C2-H10 distance, see Fig. 1) is 2.1 Å, while that for the proton transfer leading to formation of cation **D1** (C6—H10 distance) is 2.5–2.6 Å (see Table 1). This suggests that, as in the case of the QM gas-phase calculations,[10] the cation **C** will rearrange more easily to cation **F** than to cation **D1** in the TXS environment, thus, affording a pathway to the side products **V1** and **V2**. This result is in contrast with a recent study that suggests a slight preference for the direct proton transfer while employing a different model of the closed conformation of TXS.[62] This emphasizes the complexity of the TXS mechanism and the need for the fully closed conformation of TXS to be determined experimentally.
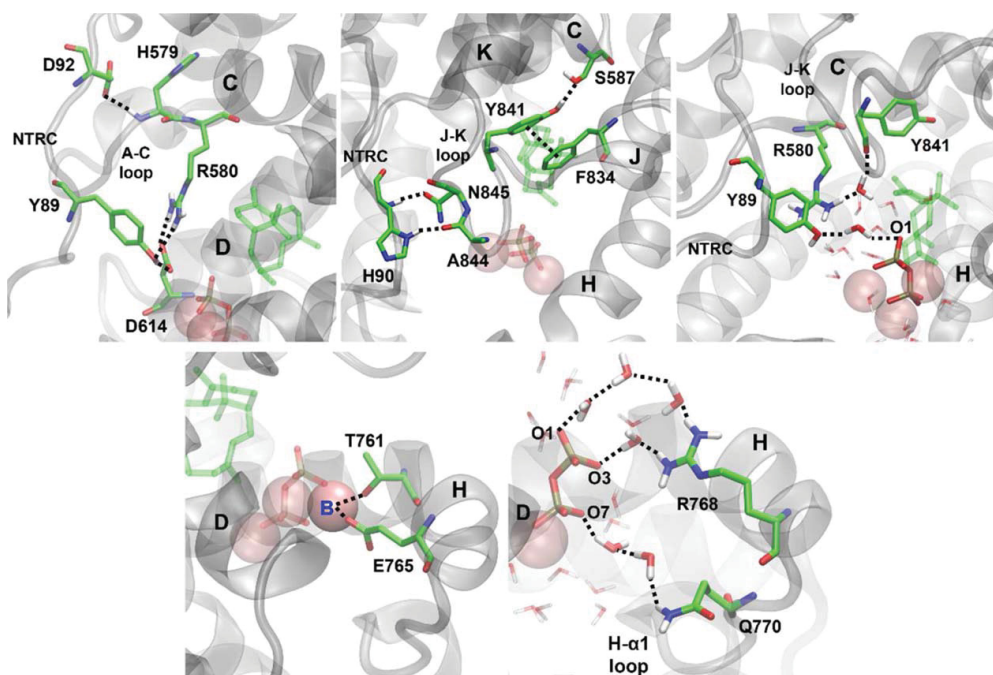
Figure 4. Representative snapshots of the **TXS·C** complex showing the interactions (dashed lines) that play a key role for the stability of the closed conformation of TXS during the MD simulations. See Table S3 in the Supporting Information for more details. [Color figure can be viewed at wileyonlinelibrary.com]

In contrast to a previous suggestion,[3] our results show that it is very unlikely for R580 to act as the base that deprotonates the **C**:C12 atom (see Fig. 1). They also indicate that though the deprotonation of **C**:C12 by PPi may be possible, it should not be the main reaction path responsible for the formation of product **V**, since the deprotonation mediated by water is expected to be favored. PROPKA[56,57] predicts a pKa value of 11.98 for the side chain of R580 at pH 7 (used in the

Figure 5. a) Superimposed structures of the **TXS·C** complex at intervals of 10 ps during MD1. MD2 and MD3 show similar superimposition profiles. b) HT-QM structure of cation **C** (purple)[10] superimposed with the best representative structure of the cluster (exhibiting the lowest RMSD with respect to the centroid of the same)[53] obtained for cation **C** from the MD simulations of the **TXS·C** complex (green). The cluster analysis employed an RMSD cutoff value of 0.60 Å, and gave only one cluster for each MD trajectory. In both (a) and (b) the hydrogen atoms of cation **C** are omitted for clarity. [Color figure can be viewed at wileyonlinelibrary.com]

experiments), which should thus be protonated at pH 7 and make this residue a poor candidate for a deprotonating base. Furthermore, in the MD simulations the average interatomic distance between the oxygen atom O1 of PPi and the closest hydrogen atom of **C**:C12 (axial hydrogen H12a, see Fig. 6) is about 4.7 Å. The shortest PPi:O1-**C**:H12a distance encountered during the MD runs is about 4.0 Å (presumably short enough to enable deprotonation of cation **C**) but it only occurs for less than 2.6% of the simulation time (see Fig. S5 in the Supporting Information). By contrast, a water molecule (W1$_C$) is frequently (overall for more than 65% of the simulation time) found very close to **C**:H12a (distance $\leq$ 3.0 Å) during the MD runs. Simultaneously, W1$_C$ forms a strong hydrogen bond (distance $\leq$ 2.0 Å) with either PPi or a second water molecule (W2$_C$), the latter in turn being strongly hydrogen bonded to either PPi or residue D614 (helix D). These water bridges may lead to the deprotonation of **C**:C12 via multiple proton transfers.[63–65] Thus, there are three possible reaction paths for a water-assisted deprotonation of cation **C**, which are referred to as DPC1, DPC2, and DPC3. DPC1 involves a double proton transfer (**C**-W1$_C$-PPi), while DPC2 and DPC3 involve a triple proton transfer (**C**-W1$_C$-W2$_C$-D614 and **C**-W1$_C$-W2$_C$-PPi, respectively; see Fig. 6). DPC1 is found to be accessible in all three MD simulations, DPC2 in two of them, and DPC3 only in a single one. This underlines the need for multiple MD simulations with different seed velocities for sampling the configurational space of enzyme-substrate complexes. Overall, structures

Table 1. Average values (in Å) of the relevant interatomic distances for the intramolecular proton transfer reactions involving the cations **C**, **D1**, and **F**, in the MD simulations of the **TXS·C**, **TXS·D1**, and **TXS·F** complexes, respectively.[a]

	Cation C		Cation F			Cation D1		
MD	C2—H10	C6—H10	Conformer (lifetime[b])	C10—H10	C6—H10	Conformer (lifetime[b])	C10—H10	C2—H10
MD1	2.1 (0.1)	2.6 (0.2)	F$_{chair}$ (61)	2.3 (0.2)	2.3 (0.2)	D1$_{chair}$ (53)	2.4 (0.2)	2.2 (0.1)
			F$_{boat}$ (39)	2.6 (0.2)	2.3 (0.2)	D1$_{boat}$ (47)	2.8 (0.2)	2.1 (0.2)
MD2	2.1 (0.1)	2.5 (0.2)	F$_{chair}$ (71)	2.4 (0.2)	2.3 (0.2)	D1$_{chair}$ (99)	2.4 (0.2)	2.1 (0.2)
			F$_{boat}$ (29)	2.8 (0.3)	2.3 (0.2)	D1$_{boat}$ (1)	2.7 (0.2)	2.0 (0.2)
MD3	2.1 (0.1)	2.5 (0.2)	F$_{chair}$ (70)	2.4 (0.2)	2.3 (0.2)	D1$_{chair}$ (100)	2.3 (0.2)	2.2 (0.2)
			F$_{boat}$ (30)	2.8 (0.2)	2.2 (0.1)			

[a] Standard deviations for average values are given in parentheses. MD1-MD3 correspond to MD simulations with different seed velocities. Two conformers are identified for cations **F** and **D1**, but only one for cation **C** (see the text). See Figure 1 for conventions on the atom labels. [b] Time of occurrence (in %) during the MD simulations.

amenable to intramolecular deprotonation reactions via DPC1-DPC3 are found to be present for 49–86% of the total simulation time (see Table S5 in the Supporting Information). All these reactions should thus be feasible and more facile than the direct deprotonation by PPi. The PPi:O1 atom acting as final proton acceptor is probably the oxygen atom that forms the ester bond in GGPP, as among the oxygen atoms of PPi O1 remains closest to **C**:C1 during the MD simulations.

Deprotonation pathway DPC1 is expected to be more favorable kinetically and thermodynamically than DPC2 and DPC3. PPi is a stronger base than an aspartate residue, which is expected to favor the proton transfers via DPC1 and DPC3 over those via DPC2. Moreover, the oxygen atom of W1$_C$ is more nucleophilic in DPC1 than in DPC3, given the direct strong hydrogen bond between W1$_C$ and PPi in the former case, which will favor the deprotonation of cation **C** via DPC1. DPC1 is thus identified to be the major contributor to the formation of product **V**. A water-assisted proton transfer to PPi has also been found responsible for the deprotonation of the camphyl carbocation during BPPS catalysis. Free energy

calculations showed this process to be favored thermodynamically by 4.8 kcal mol^{-1} and to involve a barrier of <4.0 kcal mol^{-1}.[13,23]

It should be noted that the involvement of water as a base in the catalysis of terpene synthase reactions is not a commonly accepted notion, except for enzymes generating hydroxylated products.[66] The crystal structures of terpene synthases show only one or a few water molecules to be trapped in the active site close to the carbocations, and they are usually well stabilized by their surroundings in positions where they are not well oriented to react.[13,15,66–68] Unfortunately, crystal structures of TXS complexed with analogues of the carbocation intermediates have not yet been reported, and there is no experimental information on the potential reactivity of water during TXS catalysis. However, the available crystal structure of TXS reveals that the volume of its active site is significantly larger than normally observed for terpene synthases;[5] this also holds true for the difference between this volume and that of either the substrate (GGPP) or the main product (taxadiene). Therefore, TXS may well be different
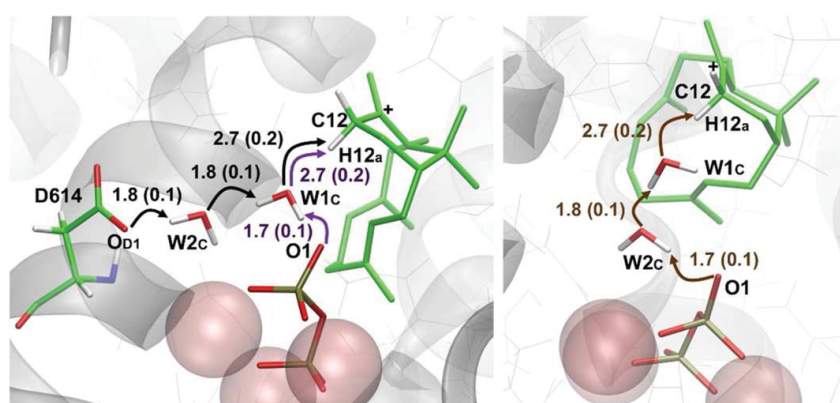

Figure 6. Possible reaction paths for the water-assisted deprotonation of cation **C** (DPC) identified in the MD simulations of the **TXS·C** complex: DPC1 (violet arrows), DPC2 (black arrows), and DPC3 (brown arrows). Average values for the relevant interatomic distances and the associated standard deviations (in parentheses) are given in Å and correspond to one of the three MD runs. See Table S5 in the Supporting Information for more details. [Color figure can be viewed at wileyonlinelibrary.com]
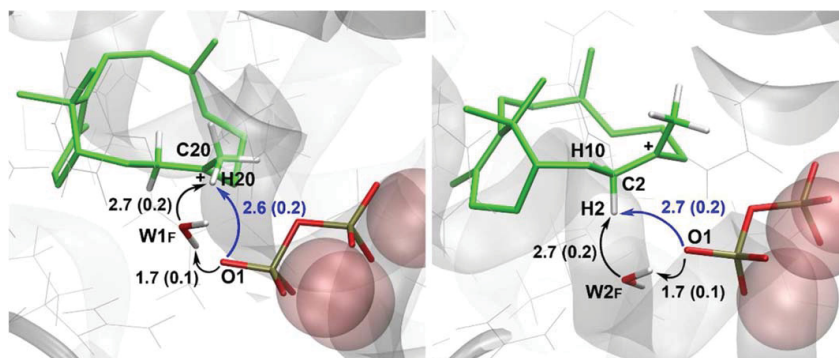
Figure 7. Possible reaction paths for the deprotonation of cation **F** (DPF) identified in the MD simulations of the **TXS·F** complex: DPF1$_{C20}$ (left, blue arrow), DPF2$_{C20}$ (left, black arrows), DPF1$_{C2}$ (right, blue arrow), and DPF2$_{C2}$ (right, black arrows). Average values for the relevant interatomic distances and the associated standard deviations (in parentheses) are given in Å and correspond to one of the three MD runs. See Table S7 in the Supporting Information for more details. [Color figure can be viewed at wileyonlinelibrary.com]

from other terpene synthases with regard to the number of water molecules that can be trapped in the active site after its closure, as well as their location and ability to react with the carbocations. It is the presence of active-site water molecules in TXS that enables the water-assisted deprotonation paths identified for cation **C** in our simulations.

TXS·F complex

The orientation of cation **F** inside the active site of TXS is also very stable throughout the MD simulations of the **TXS·F** complex (see Fig. S6 and Table S6 in the Supporting Information). Two conformers of cation **F** are found along the trajectories, as the A ring flips between a chair-like (**F$_{chair}$**) and a boat-like (**F$_{boat}$**) conformation. **F$_{chair}$** is equivalent to the HT-QM structure and is present over 60% of the simulation time, compared to less than 40% for **F$_{boat}$** (see Table 1 and Fig. S7 in the Supporting Information). The average interatomic distance for the backward intramolecular proton transfer connecting cations **C** and **F** (C10—H10 distance, see Fig. 1) is 0.3–0.4 Å larger for **F$_{boat}$** in comparison to **F$_{chair}$**, while that for the proton transfer leading to formation of cation **D1** (C6-H10 distance) is similar for both conformers (see Table 1). Nevertheless, the backward reaction to cation **C** seems possible in **F$_{boat}$**, as C10—H10$_{boat}$≤2.8 Å, leading to cation **C** with the A ring in a boat-like conformation. This suggests the existence of a variant of the HT-QM mechanism for taxadiene formation, involving preorganization of the terminal diene of GGPP to a boat conformer. In this pathway (hereafter referred to as GGPP$_{boat}$ pathway), the transformation of cation **F** to cation **D1** could have a lower energy barrier in comparison to the backward reaction to cation **C**, as C6—H10$_{boat}$ is on average 0.3–0.6 Å shorter than C10—H10$_{boat}$. By contrast, the barriers for these two reactions are expected to be not very different in the HT-QM mechanism, as the average distances C6—H10$_{chair}$ and C10—H10$_{chair}$ are similar. Hence, the HT-QM and GGPP$_{boat}$ pathways should have different energy requirements, which could be checked by computing their energy profiles in the

TXS environment to fully understand TXS catalysis (see also the next sections).[33,41,69] It should also be noted that the A ring in taxadiene adopts a boat-like conformation.[18] As shown in Figure 1, in the HT-QM pathway the A ring changes from a chair-like to a boat-like conformation during the transformation of cation **D2** to cation **E**.[10] This conversion of the A ring would not be needed on the GGPP$_{boat}$ pathway.

Regarding deprotonation of cation **F**, we found two reaction paths for deprotonation involving either the **F**:C20 (formation of product **V1**) or **F**:C2 (formation of product **V2**) atom (see Fig. 1). These pathways correspond to a direct and a water-assisted proton transfer to PPi, which are referred to as DPF1$_{C20/C2}$ and DPF2$_{C20/C2}$, respectively (see Fig. 7). In the MD simulations, structures that can be assigned to the DPF1$_{C20}$ and DPF2$_{C20}$ regimes are found to be present during 30–70% and 3–27% of the total simulation time, respectively. Furthermore, the average interatomic distance relevant for the direct deprotonation by PPi (**F**:H20-PPi:O1) is 0.1–0.2 Å shorter than that relevant for the water-mediated deprotonation (**F**:H20-W1$_F$:O). Considering these results and the stronger basicity of PPi, deprotonation is expected to be more facile via DPF1$_{C20}$ than via DPF2$_{C20}$. Hence, DPF1$_{C20}$ should represent the main reaction path responsible for formation of **V1**.

By contrast, structures that can be assigned to the DPF2$_{C2}$ regime are encountered very frequently in the MD runs (over 70% of the simulation time) while those in the DPF1$_{C2}$ regime are found only rarely (less than 7% of the simulation time). This suggests that the DPF2$_{C2}$ path is mainly responsible for formation of **V2**, even though DPF1$_{C2}$ may benefit from the direct participation of PPi in the deprotonation of the cation.[36,38] Our results are consistent with the experimentally observed higher yield of **V2** (compared with **V1**), given the higher population of structures in the DPF2$_{C2}$ regime (compared with DPF1$_{C20}$, see above) and the obstacles to the deprotonation of cation **D** (see next section). In contrast to previous notions,[3] the MD simulations of the **TXS·F** complex thus suggest that the formation of side products **V1** and **V2** mainly proceeds through deprotonation of cation **F** on two
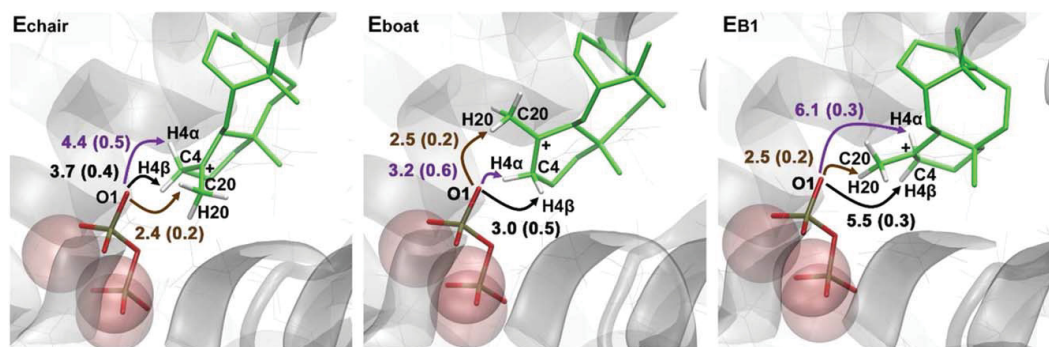
79

Figure 8. Conformers of cation **E** identified in the MD simulations of the **TXS·E** complex. Average values for the relevant interatomic distances and the associated standard deviations (in parentheses) are given in Å. For each conformer, averages were calculated over all MD runs in which this conformer was encountered (see the text). [Color figure can be viewed at wileyonlinelibrary.com]

different pathways. See Table S7 in the Supporting Information for more details on the DPFs.

Conversely, our MD simulations are in line with the previous finding that PPi plays an important role in the thermodynamics of the reaction.[3] The MD simulations yield an average PPi:O1-**F**:C3(+) distance of about 4.0 Å in the **TXS·F** complex, compared with an average PPi:O1-**C**:C11(+) distance of about 6.5 Å in the **TXS·C** complex. PPi will, therefore, help stabilize **TXS·F** relative to **TXS·C** through electrostatic interactions. The formation of **TXS·F** from **TXS·C** will, thus, be facilitated thermodynamically in the enzyme, while it is endothermic in the gas phase.[3,10]

TXS·D1 complex

As in the case of cation **F**, the orientation of cation **D1** in the active site is very stable during the MD simulations of the **TXS·D1** complex (see Table S8 and Fig. S8 in the Supporting Information). There are two conformers of the cation that are interconverted by the flipping of the A ring between a chair-like (**D1$_{chair}$**, equivalent to the HT-QM structure) and boat-like (**D1$_{boat}$**) conformation. The accessibility of **D1$_{boat}$** in these MD simulations supports the feasibility of the GGPP$_{boat}$ pathway. Both **D1$_{chair}$** and **D1$_{boat}$** are unproductive toward subsequent cyclization to cation **E** (C2—C7 σ-bond formation), since they both adopt the same unfavorable conformation of the 12-membered ring, with an average interatomic C2—C7 distance of >4.0 Å in the MD simulations.[10] As shown in Figure 1, cation **D1** must undergo a conformational change to bring the positively charged C7 atom closer to the C2=C3 π-bond to generate a productive conformer (cation **D2**) that can rearrange to cation **E**. This process should be possible in the TXS environment both on the HT-QM and GGPP$_{boat}$ pathways. The energy barrier for this conformational change is about 4 kcal mol^{-1} in the gas phase[10] and may well be lower in the TXS enzyme. Meanwhile, the average C10—H10 distance is larger than the average C2—H10 distance for both **D1$_{boat}$** (by 0.7 Å) and **D1$_{chair}$** (by 0.1-0.3 Å), suggesting that the reaction **D1** → **C** is less facile than the reaction **D1** → **F** on both pathways;

this propensity should be more pronounced for **D1$_{boat}$** given the larger difference between the distances (see Table 1).

The MD simulations of the **TXS·D1** complex also show that cation **D1** is not involved in the formation of the product **V2**. The atoms **D1**:C6 (i.e., the expected deprotonation target) and **D1**:C7(+) remain oriented toward the interior of the catalytic cavity, far away from PPi, during all MD simulations (see Fig. S8 in the Supporting Information). The average interatomic distances between these carbon atoms and PPi:O1 are about 7.0 and 8.0 Å, respectively. Furthermore, there are no surrounding protein residues that could act as a base and deprotonate **D1**:C6.

The large PPi:O1-**D**:C7(+) distance implies that the electrostatic interactions between PPi and the cationic center contribute little to the stabilization of **TXS·D1**, certainly less than in the case of **TXS·C** and especially **TXS·F** (see previous section). PPi, thus, appears to favor the formation of cation **D1** via the two-step route (**C**-to-**F**-to-**D1**) rather than the one-step route (**C**-to-**D1**; see Fig. 1). In addition to participating in the deprotonation of cation **F**, PPi also mediates the formation of the side products **V1** and **V2** by electrostatically guiding the reaction through the formation of cation **F**.[23] This is true for both the HT-QM and GGPP$_{boat}$ pathways, as it is related to the orientation of the cation in the active site of TXS, which is very stable during the MD simulations of the complexes in question as described above.

TXS·E complex

Three different conformers of cation **E** are found in the MD simulations of the **TXS·E** complex which are labeled as **E$_{chair}$**, **E$_{t-boat}$**, and **E$_{B1}$** (see Fig. 8). **E$_{chair}$**, which is equivalent to the (starting) HT-QM structure, is present along the whole trajectory in one of the three MD runs. In the other two runs, **E$_{chair}$** is converted to **E$_{t-boat}$** (after 291 ps of production) and **E$_{B1}$** (during equilibration). **E$_{t-boat}$** differs from **E$_{chair}$** with regard to the C ring, which has a twist boat-like conformation in the former and a chair-like conformation in the latter. Meanwhile, **E$_{B1}$** differs from **E$_{chair}$** in the conformation of the B ring as well as

the orientation of the C ring with respect to the A and B rings (see Fig. S9 in the Supporting Information). These conformational changes mainly affect the interaction between the C ring and PPi, as the orientation of the cation as a whole in the active site remains overall the same throughout all MD simulations (see Fig. 8 and Table S9 and Fig. S10 of the Supporting Information).

The only possible deprotonation mechanism for cation **E** is the direct proton transfer to PPi. As indicated by the average interatomic distances between PPi and the target hydrogen atoms for deprotonation (see Fig. 8), all three conformers of cation **E** can easily undergo deprotonation at C20, while deprotonation at C4 is by far favored in the $E_{t\text{-boat}}$ conformation. Regarding the latter, deprotonation on the β face of C4 is expected to predominate over that on the α face, as the average PPi:O1-H4β distance is shorter than PPi:O1-H4α, and the twist boat conformation of the C ring also favors elimination of H4β stereoelectronically. This is in agreement with labeling experiments using deuterated GGPP.[17,18] The average distance between PPi and the positively charged carbon atom of cation **E** is very similar for E_{chair} and $E_{t\text{-boat}}$ (3.8 and 3.7 Å), and much shorter than in the case of cations **C** and **F** (see above). This suggests that the **TXS·E** complexes are lower in energy than the **TXS·C** and **TXS·F** complexes, which correlates well with the experimentally observed yields: the taxadiene products **T**-**T1** (arising from **TXS·E**) dominate over the side products **V**-**V2** (arising from the other complexes, see Fig. 1).

Conversely, the predominance of **T** over **T1** in the yield cannot be rationalized from the average interatomic distances between PPi and the hydrogen atoms bound to C4 and C20 of cation **E**, respectively. For all three conformers of **TXS·E** (i.e., with E_{chair}, $E_{t\text{-boat}}$, or E_{B1}) the computed distances favor the formation of **T1** rather than that of **T**. To resolve this problem, it is necessary to go beyond MD simulations of the different **TXS·C**-**TXS·F** complexes and to compute the energy profiles for the whole reaction scheme (for both the HT-QM and GGPP$_{boat}$ variants, and including the deprotonation reactions of cations **C** and **F** leading to the side products **V**-**V2**). We are currently performing such a study at the QM/MM level.[33,40–42,69,70] The results will be reported in a separate paper.

Conclusions

We have performed MD simulations of enzyme carbocation complexes (**TXS·C**, **TXS·F**, **TXS·D1**, and **TXS·E**), which are essential reaction intermediates in the cyclization of GGPP to taxadiene (**T**) catalyzed by TXS. These simulations provide insight into the promiscuity of TXS by identifying possible pathways to the four observed side products (**V**, **V1**, **V2**, and **T1**, see Fig. 1). In agreement with a previous static study (employing docking and MM optimizations),[3] we find that PPi plays a key role in TXS catalysis. PPi favors **TXS·E** thermodynamically over the other complexes, consistent with the experimentally observed high yield of taxadiene that is formed by deprotonation of **TXS·E**. The MD simulations do not support the hypothesis that carbocation tumbling enables deprotonation of all carbocations

by PPi, thus, leading to the formation of all observed products. The cavity of the active site of TXS harboring the carbocations during the catalytic process is overall rather rigid and restricts the positional freedom of the carbocations.

Judging from the average values of the catalytically relevant interatomic distances in the MD simulations, an exclusive deprotonation of the carbocation by PPi is only possible in the **TXS·E** complex. Regarding **TXS·F**, cation **F** can be deprotonated either directly by PPi or by a water-assisted proton transfer to PPi to produce either **V1** or **V2**. In the absence of carbocation tumbling, it does not seem feasible to deprotonate cation **D1** in the **TXS·D1** complex. Finally, in the **TXS·C** complex, cation **C** is expected to be deprotonated to yield **V** via multiple proton transfer reactions assisted by water bridges, with either PPi or an aspartate residue at helix D (D614) acting as the final proton acceptor.

The MD simulations of the **TXS·cation** complexes indicate that formation of **T** in the TXS environment via the HT-QM pathway requires a conformational change of the C ring of cation **E** from a chair-like to a boat-like conformation. In the resulting **TXS·E** complex, cation **E** is in a much better orientation to be deprotonated by PPi at C4. This conformational change of the C ring correlates well with the results of labeling experiments using deuterated GGPP.[17,18]

Conversely, the MD simulations also suggest a variant of the HT-QM pathway that would account for the formation of taxadiene and the side products. This pathway (GGPP$_{boat}$) involves preorganization of the terminal diene of GGPP in a boat folding at the initial stage of the reaction. Analysis of the catalytically relevant interatomic distances during the MD runs indicates that the GGPP$_{boat}$ and HT-QM pathways will be energetically different. The present MD simulations do not provide a definitive assessment which of these two pathways is more favorable. This requires the computation of complete energy profiles in the TXS environment for the reactions considered presently (conversion of GGPP to **T** and formation of side products), covering not only the intermediates (**TXS·cation**) as in the MD runs but also the relevant transition states and reaction paths. We are currently performing a corresponding QM/MM study and will report the results separately.

[1] M. Hezari, N. G. Lewis, R. Croteau, *Arch. Biochem. Biophys.* **1995**, *322*, 437.

[2] D. C. Williams, M. R. Wildung, A. Q. Jin, D. Dalal, J. S. Oliver, R. M. Coates, R. Croteau, *Arch. Biochem. Biophys.* **2000**, *379*, 137.

A. Paper I

[3] P. Schrepfer, A. Buettner, C. Goerner, M. Hertel, J. van Rijn, F. Wallrapp, W. Eisenreich, V. Sieber, R. Kourist, T. Brück, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E958.

[4] X. Lin, M. Hezari, A. E. Koepp, H. G. Floss, R. Croteau, *Biochemistry* **1996**, *35*, 2968.

[5] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson, *Nature* **2011**, *469*, 116.

[6] R. N. Patel, *Annu. Rev. Microbiol.* **1998**, *52*, 361.

[7] G. M. Cragg, *Med. Res. Rev.* **1998**, *18*, 315.

[8] S. Howat, B. Park, I. S. Oh, Y.-W. Jin, E.-K. Lee, G. J. Loake, *New Biotechnol.* **2014**, *31*, 242.

[9] S. Malik, R. M. Cusidó, M. H. Mirjalili, E. Moyano, J. Palazón, M. Bonfill, *Process Biochem.* **2011**, *46*, 23.

[10] Y. J. Hong, D. J. Tantillo, *J. Am. Chem. Soc.* **2011**, *133*, 18249.

[11] S. Soliman, Y. Tang, *Biotechnol. Bioeng.* **2015**, *112*, 229.

[12] M. R. Wildung, R. Croteau, *J. Biol. Chem.* **1996**, *271*, 9201.

[13] D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau, D. W. Christianson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15375.

[14] L. C. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini, *Biochemistry* **1994**, *33*, 10871.

[15] D. W. Christianson, *Chem. Rev.* **2006**, *106*, 3412.

[16] P. Gutta, D. J. Tantillo, *Org. Lett.* **2007**, *9*, 1069.

[17] Y. Jin, D. C. Williams, R. Croteau, R. M. Coates, *J. Am. Chem. Soc.* **2005**, *127*, 7834.

[18] Q. Jin, D. C. Williams, M. Hezari, R. Croteau, R. M. Coates, *J. Org. Chem.* **2005**, *70*, 4667.

[19] S. Y. Chow, H. J. Williams, Q. Huang, S. Nanda, A. I. Scott, *J. Org. Chem.* **2005**, *70*, 9997.

[20] S. Y. Chow, H. J. Williams, J. D. Pennington, S. Nanda, J. H. Reibenspies, A. I. Scott, *Tetrahedron* **2007**, *63*, 6204.

[21] D. C. Williams, B. J. Carroll, Q. Jin, C. D. Rithner, S. R. Lenger, H. G. Floss, R. M. Coates, R. M. Williams, R. Croteau, *Chem. Biol.* **2000**, *7*, 969.

[22] S. Edgar, F.-S. Li, K. Qiao, J.-K. Weng, G. Stephanopoulos, *ACS Synth. Biol.* **2017**, *6*, 201.

[23] D. T. Major, M. Weitman, *J. Am. Chem. Soc.* **2012**, *134*, 19454.

[24] D. W. Christianson, *Curr. Opin. Chem. Biol.* **2008**, *12*, 141.

[25] M. W. van der Kamp, J. Sirirak, J. Żurek, R. K. Allemann, A. J. Mulholland, *Biochemistry* **2013**, *52*, 8094.

[26] C. M. Starks, K. Back, J. Chappell, J. P. Noel, *Science* **1997**, *277*, 1815.

[27] K. Zhou, R. J. Peters, *Chem. Commun.* **2011**, *47*, 4074.

[28] E. Y. Shishova, F. Yu, D. J. Miller, J. A. Faraldos, Y. Zhao, R. M. Coates, R. K. Allemann, D. E. Cane, D. W. Christianson, *J. Biol. Chem.* **2008**, *283*, 15431.

[29] Y. J. Hong, D. J. Tantillo, *Org. Biomol. Chem.* **2010**, *8*, 4589.

[30] P. Trodler, R. D. Schmid, J. Pleiss, *BMC Struct. Biol.* **2009**, *9*, 38.

[31] P. Trodler, J. Pleiss, *BMC Struct. Biol.* **2008**, *8*, 9.

[32] A. M. Escorcia, M. C. Daza, M. Doerr, *J. Mol. Catal. B Enzym.* **2014**, *108*, 21.

[33] A. M. Escorcia, K. Sen, M. C. Daza, M. Doerr, W. Thiel, *ACS Catal.* **2017**, *7*, 115.

[34] L. S. D. Caves, J. D. Evanseck, M. Karplus, *Protein Sci.* **1998**, *7*, 649.

[35] G. A. Worth, F. Nardi, R. C. Wade, *J. Phys. Chem. B* **1998**, *102*, 6260.

[36] K. Arora, C. L. Brooks III, In *Dynamics in Enzyme Catalysis*; J. Klinman, S. Hammes-Schiffer, Eds.; Springer: Berlin, Heidelberg, **2013**; pp. 165–187.

[37] G. G. Hammes, S. J. Benkovic, S. Hammes-Schiffer, *Biochemistry* **2011**, *50*, 10422.

[38] P.-O. Syrén, K. Hult, *Chembiochem* **2010**, *11*, 802.

[39] Y. Zhang, J. Kua, J. A. McCammon, *J. Phys. Chem. B* **2003**, *107*, 4459.

[40] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.

[41] B. Karasulu, M. Patil, W. Thiel, *J. Am. Chem. Soc.* **2013**, *135*, 13400.

[42] K. Sen, W. Thiel, *J. Phys. Chem. B* **2014**, *118*, 2810.

[43] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.

[44] P. H. Hünenberger, In *Advanced Computer Simulation*; C. D. Holm, K. P. D. Kremer, Eds.; Springer: Berlin, Heidelberg, **2005**; pp. 105–149.

[45] E. Sanchez-Garcia, M. Doerr, Y.-W. Hsiao, W. Thiel, *J. Phys. Chem. B* **2009**, *113*, 16622.

[46] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.

[47] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260.

[48] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.

[49] A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400.

[50] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, M. Karplus, *J. Phys. Chem. B* **2001**, *105*, 569.

[51] R. H. Stote, D. J. States, M. Karplus, *J. Chim. Phys.* **1991**, *88*, 2419.

[52] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.

[53] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem. Int. Ed.* **1999**, *38*, 236.

[54] B. Christelle, B. D. O. Eduardo, C. Latifa, M. Elaine-Rose, M. Bernard, R.-H. Evelyne, G. Mohamed, E. Jean-Marc, H. Catherine, *J. Biotechnol.* **2011**, *156*, 203.

[55] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *Nucleic Acids Res.* **2004**, *32*, W665.

[56] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525.

[57] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 2284.

[58] E. Y. Shishova, L. Di Costanzo, D. E. Cane, D. W. Christianson, *Biochemistry* **2007**, *46*, 1941.

[59] P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2014**, *53*, 7652.

[60] T. Tokiwano, T. Endo, T. Tsukagoshi, H. Goto, E. Fukushi, H. Oikawa, *Org. Biomol. Chem.* **2005**, *3*, 2713.

[61] M. J. Begley, C. B. Jackson, G. Pattenden, *Tetrahedron* **1990**, *46*, 4907.

[62] Y. Freud, T. Ansbacher, D. T. Major, *ACS Catal.* **2017**, *7*, 7653.

[63] Z. Smedarchina, W. Siebrand, A. Fernández-Ramos, Q. Cui, *J. Am. Chem. Soc.* **2003**, *125*, 243.

[64] B. K. Mai, K. Park, M. P. T. Duong, Y. Kim, *J. Phys. Chem. B* **2013**, *117*, 307.

[65] D. Riccardi, P. König, X. Prat-Resina, H. Yu, M. Elstner, T. Frauenheim, Q. Cui, *J. Am. Chem. Soc.* **2006**, *128*, 16302.

[66] D. W. Christianson, *Chem. Rev.* **2017**, *117*, 11570.

[67] M. Chen, W. K. W. Chou, N. Al-Lami, J. A. Faraldos, R. K. Allemann, D. E. Cane, D. W. Christianson, *Biochemistry* **2016**, *55*, 2864.

[68] Y. Gao, R. B. Honzatko, R. J. Peters, *Nat. Prod. Rep.* **2012**, *29*, 1153.

[69] S. Metz, W. Thiel, *J. Am. Chem. Soc.* **2009**, *131*, 14885.

[70] I. Polyak, M. T. Reetz, W. Thiel, *J. Am. Chem. Soc.* **2012**, *134*, 2732.

82

Supporting Information

Molecular dynamics study of taxadiene synthase catalysis

*Andrés M. Escorcia, Jeaphianne P. M. van Rijn, Gui-Juan Cheng, Patrick Schrepfer,
Thomas B. Brück, and Walter Thiel**

[*]W.T.: Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim, Germany.
Email: thiel@kofo.mpg.de

Content:

1

1. Additional details on the MD setup

The residues kept fixed during the MD simulations correspond to residues of the N-terminal domain that are distant from the reaction core and to residues at the surface of the C-terminal domain (Figure S1). These residues are listed below:

➢ N-terminal: residues 101-130 and 537-552.

➢ C-terminal: residues 553-554, 592-594, 634, 636, 638-639, 641-642, 666-672, 675-676, 697, 699, 702, 704, 728-734, 738, 742, 778-789, 792-793, 796, 799, 803, 807, 809-818, 821 and 857-862.
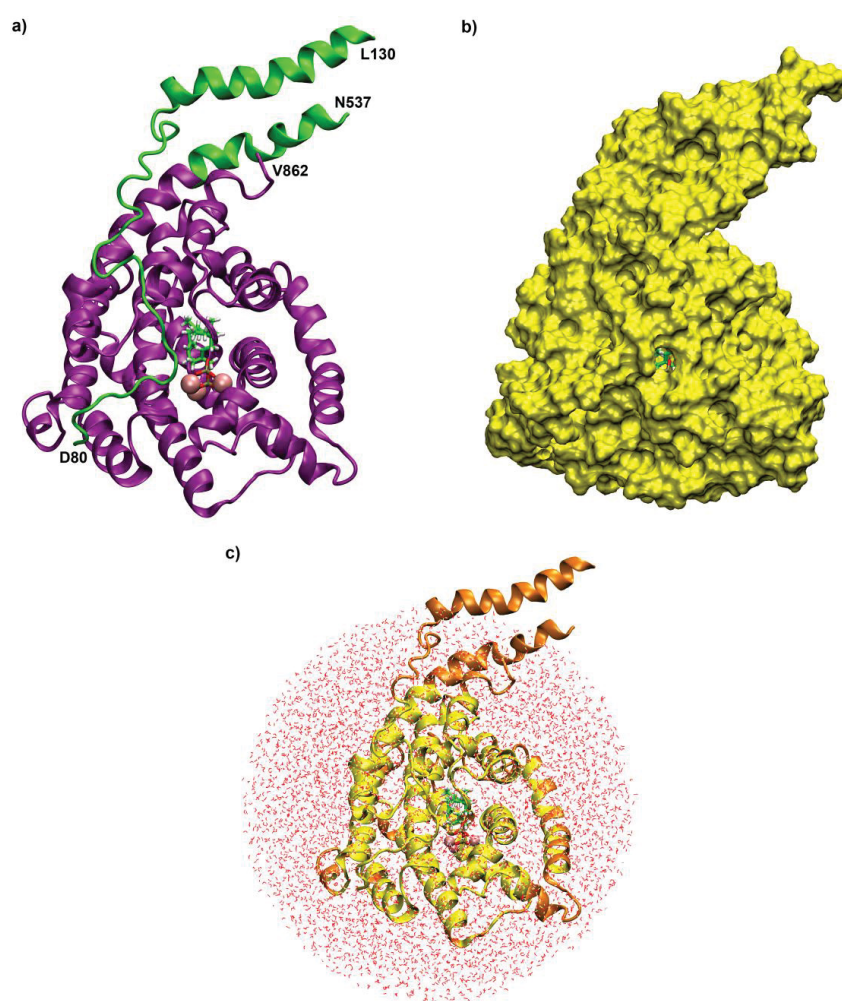
Figure S1. (a) Representative structure of TXS used in the MD simulations. It comprises the C-terminal domain (S553-V862, purple) and part of the N-terminal domain (D80-L130 and N537-Q552, green). (b) Surface representation of TXS showing more clearly that the active site is enclosed. (c) Solvated system with the protein residues kept fixed during the MD simulations depicted in orange.

2

2. Structural stability of the protein

Table S1

Average RMSD (with respect to the initial structure) for all heavy atoms of the protein backbone in the MD simulations of the **TXS·cation** complexes[a]

COMPLEX	RMSD (Å)		
	MD1	MD2	MD3
TXS·C	1.2 (0.2)	1.04 (0.07)	1.07 (0.08)
TXS·D1	1.1 (0.1)	0.97 (0.05)	1.2 (0.2)
TXS·F	1.1 (0.1)	0.97 (0.08)	1.07 (0.04)
TXS·E	1.1 (0.1)	1.04 (0.05)	1.03 (0.05)

[a] Standard deviations are given in parentheses. MD1-MD3 correspond to MD simulations with different initial velocity distributions.

Table S2

Average RMSD (with respect to the initial structure) for all atoms of PPi in the MD simulations of the **TXS·cation** complexes[a]

COMPLEX	RMSD (Å)		
	MD1	MD2	MD3
TXS·C	0.8 (0.2)	0.7 (0.1)	0.8 (0.2)
TXS·D1	0.5 (0.2)	0.5 (0.2)	0.4 (0.1)
TXS·F	0.6 (0.2)	0.5 (0.2)	0.6 (0.2)
TXS·E	0.5 (0.1)	0.5 (0.2)	0.5 (0.1)

[a] Standard deviations are given in parentheses. MD1-MD3 correspond to MD simulations with different initial velocity distributions. The RMSD values for PPi are lower than those for the protein, which is indicative of its high positional and conformational stability in the active site of TXS.

3

Table S3
Interactions identified to play a key role for the stability of the closed conformation of TXS (indicated by CLO) and for the rigidity of the cavity of the active site harboring the carbocation (indicated by RIG) in the MD simulations of the **TXS·cation** complexes

Interactions[a]	Role		Lifetime[d]			
	RIG	CLO	C	D1	F	E
hydrogen bonds[b]						
PPi:O1,O3,O4-N757:*ND2*	X		97	99	97	99
PPi:O5,O7-R754:*NH1,NH2*	X		100	100	100	100
D614:OD1,OD2-R580:*NH1,NH2*	X		100	62	48	65
D614:OD1,OD2-Y89:*OH*		X	95	97	84	67
Y89:OH-R580:*NH1,NH2*	X	X	83	72	56	68
S587:OG-Y841:*OH*	X	X	99	94	95	98
N845:OD1-H90:*N,ND1*		X	85	70	64	100
A844:O-H90:*ND1*		X	43	66	21	47
D92:OD1,OD2-H579:*N*		X	100	96	100	100
G715:O-Q609:*NE2*	X		92	100	93	69
Y684:<u>OH</u>-Y688:<u>OH</u>	X		99	100	100	89
Y835:OH-N757:*ND2*	X		92	68	91	68
water-mediated hydrogen bonds[b,c]						
PPi:O1-W$_{(1,2)}$-Y835:*OH*	X		100	100	100	100
PPi:O1-W$_{(1)}$-Y89:<u>OH</u>		X	99	65	54	70
PPi:O6,O7-W$_{(2-3)}$-Q770:OE1,*NE2*		X	100	100	100	100
PPi:O1-W$_{(1-3)}$-R580:*NH2*	X		77	100	100	91
PPi:O1,O2,O3-W$_{(1-3)}$-R768:*NH1,NH2*		X	100	100	100	100
Y841:O-W$_{(1)}$-R580:*NH2*	X	X	66	37	31	62
S713:OG-W$_{(1-2)}$-Mg^{2+}$_C$	X		88	60	31	85
metal-ligand coordinations						
N757:OD1-Mg^{2+}$_B$	X		100	100	100	100
T761:OG1-Mg^{2+}$_B$		X	100	100	100	100
E765:OE1,OE2-Mg^{2+}$_B$		X	100	100	100	100
D613:OD2-Mg^{2+}$_A$	X		100	100	100	100
D613:OD1-Mg^{2+}$_C$	X		100	100	100	100
π-stacking Y841-F834	X	X	100	100	100	100

[a] There are pairs of residues that form different hydrogen bonds during the MD simulations (e.g. D614 and R580), some of which occur simultaneously during certain intervals of time. In this case the respective residues are specified once, and all acceptor and/or donor atoms are listed separated by commas. The same notation is used for residues coordinating to a Mg^{2+} ion through more than one atom.

[b] Atoms acting as hydrogen bond acceptors and donors are indicated in regular and italic fonts, respectively, whereas atoms changing from donor to acceptor character and vice versa during the MD simulations are underlined. A maximum distance of 3 Å between acceptor and hydrogen was chosen as criterion for the presence of a hydrogen bond.

[c] W$_{(n)}$ refers to water molecules, "n" being their number. "n" is given as a range when the number of water molecules fluctuates during and/or between the MD simulations.

[d] The entry under Lifetime is the percentage of the simulation time that the type of interaction between the respective pair of residues (independent of the atoms involved) occurs during the MD simulations. The values reported here correspond to average values obtained from the three MD runs with different seed velocities.

4
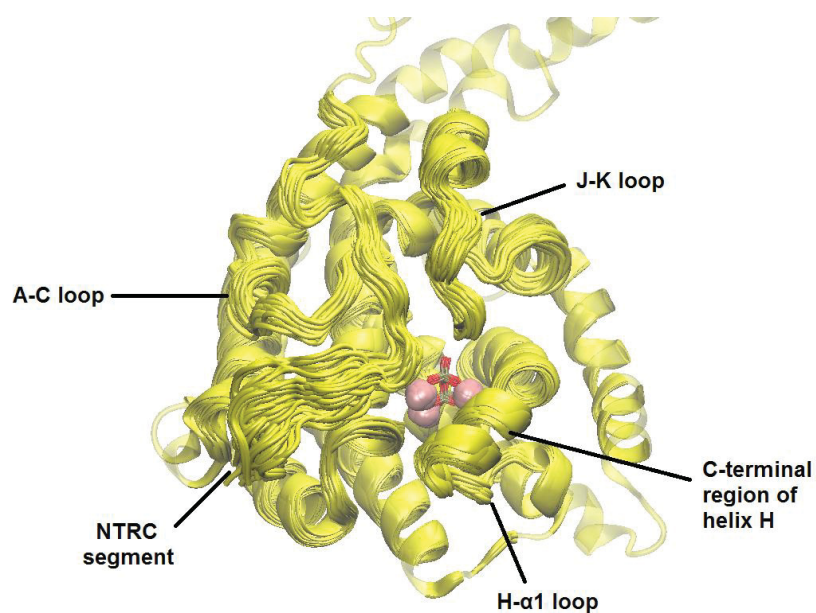
Figure S2. Superimposed structures of PPi (oxygens in red), catalytic Mg^{2+} ions (pink) and TXS (yellow), at intervals of 10 ps during one of the three MD runs of the **TXS·C** complex. All other simulations show similar superimposition profiles. Protein residues which were kept fixed during the MD simulations are shown as transparent loops (on top, see also Figure S1c).
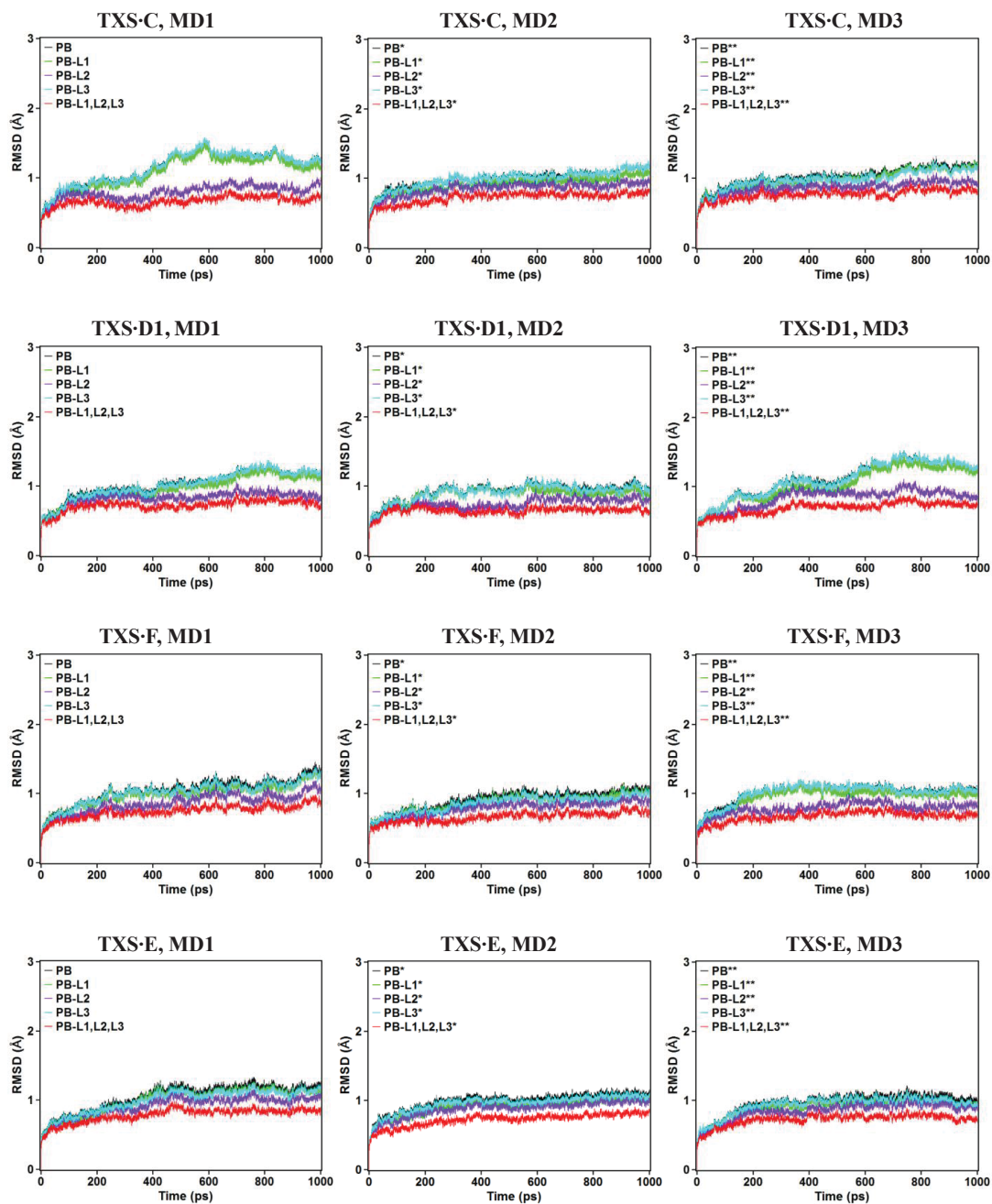
5

Figure S3. Time evolution of the RMSD for the protein structure in the MD simulations of the **TXS·cation** complexes. PB stands for all heavy atoms of the protein backbone, while "PB-L1, L2 and/or L3" refers to PB without the A-C loop (L1), the NTRC segment (L2) and/or the J-K loop (L3). MD1-MD3 correspond to MD simulations with different seed velocities. Superscripts * are added to distinguish between these MD simulations for more clarity. In all MD simulations the RMSD values become smaller overall when removing the contribution of these protein regions, especially that from the NTRC segment, which is indicative of its high flexibility.

6
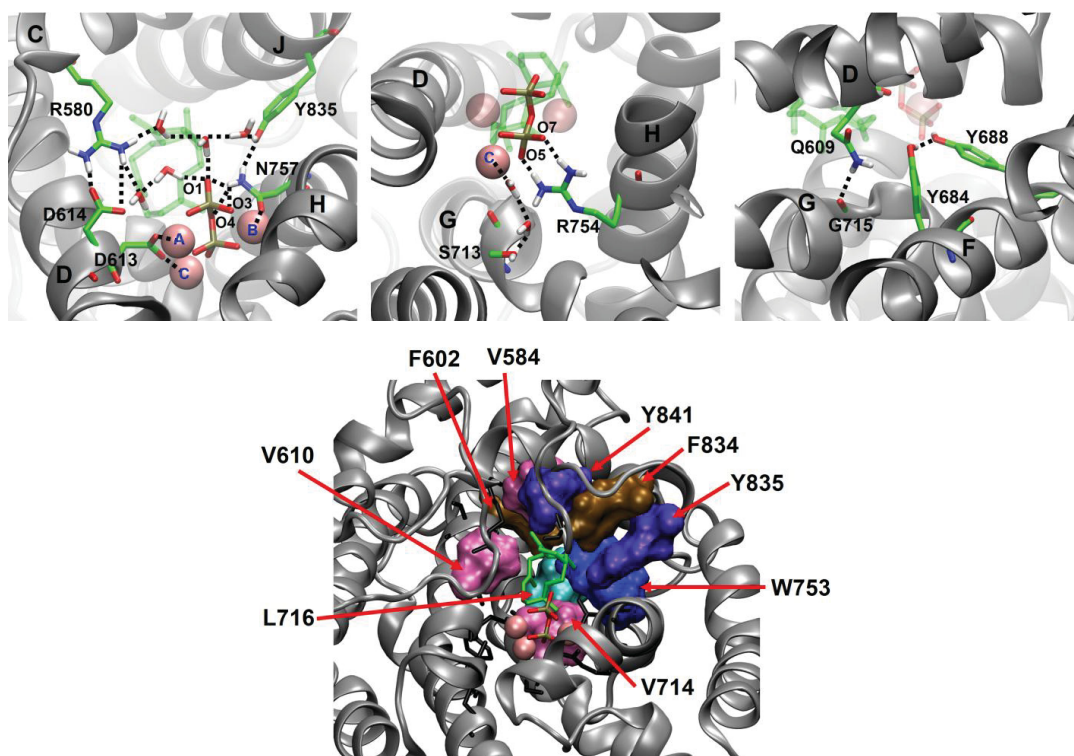
Figure S4. Representative snapshots of the **TXS·C** complex to illustrate the interactions (top) and the residues with bulky side chains (bottom) which contribute to the rigidity of the active-site cavity of TXS harboring the carbocations. Interactions that are important for both the rigidity of the cavity and the stability of the closed conformation of TXS (see Table S3) are omitted, as they are shown in Figure 4 of the main text.

3. Cation C

Table S4

Average RMSD (with respect to the initial structure) for all carbon atoms of cation **C** in the MD simulations of the **TXS·C** complex[a]

MD	RMSD (Å)
MD1	0.8 (0.2)
MD2	0.9 (0.4)
MD3	0.8 (0.2)

[a] Standard deviations are given in parentheses. MD1-MD3 correspond to MD simulations with different initial velocity distributions. The low RMSD values and standard deviations in all simulations are indicative of the high positional and conformational stability of cation **C** in the active site of TXS.
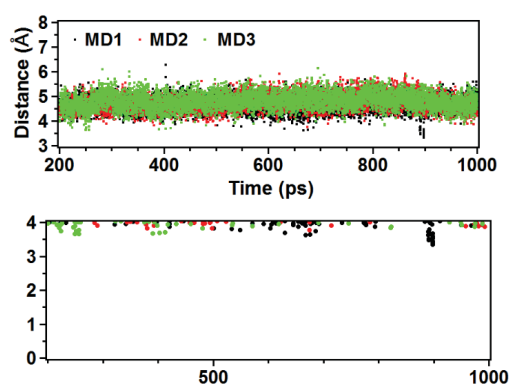
Figure S5. Time evolution of the PPi:O1-**C**:H12a distance in the MD simulations of the **TXS·C** complex. The periods of time when PPi:O1-**C**:H12a ≤ 4.0 Å are highlighted in the panel at the bottom. MD1-MD3 correspond to MD simulations with different initial velocity distributions. See Figure 6 of the main text for conventions on the atom labels.

8

Table S5

Average values of the relevant interatomic distances and angles involved in the possible water-assisted deprotonation paths of cation **C** (DPC) identified in the MD simulations of the **TXS·C** complex[a]

DPC (lifetime[b], %)	Distance ; Angle Definitions	Distance (Å) ; Angle (°) Values		
		MD1	**MD2**	**MD3**
DPC1 (86 - 61 - 9)	**C**:H12a-$W1_C$:O ; **C**:(C12-H12a)-$W1_C$:O	2.7 (0.2) ; 142 (12)	2.7 (0.2) ; 132 (12)	2.8 (0.2) ; 150 (11)
	$W1_C$:H-PPi:O1 ; $W1_C$:(O-H)-PPi:O1	1.7 (0.1) ; 165 (8)	1.7 (0.1) ; 165 (8)	1.7 (0.1) ; 162 (9)
DPC2 (65 - 49 - 0)	**C**:H12a-$W1_C$:O ; **C**:(C12-H12a)-$W1_C$:O	2.7 (0.2) ; 142 (11)	2.7 (0.2) ; 132 (13)	NI[c]
	$W1_C$:H-$W2_C$:O ; $W1_C$:(O-H)-$W2_C$:O	1.8 (0.1) ; 165 (8)	1.8 (0.1) ; 164 (8)	
	$W2_C$:H-D614:O_{D1} ; $W2_C$:(O-H)-D614:O_{D1}	1.8 (0.1) ; 161 (10)	1.8 (0.1) ; 161 (10)	
DPC3 (0 - 0 - 74)	**C**:H12a-$W1_C$:O ; **C**:(C12-H12a)-$W1_C$:O	NI[c]	NI[c]	2.7 (0.2) ; 127 (12)
	$W1_C$:H-$W2_C$:O ; $W1_C$:(O-H)-$W2_C$:O			1.8 (0.1) ; 162 (9)
	$W2_C$:H-PPi:O1 ; $W2_C$:(O-H)-PPi:O1			1.7 (0.1) ; 162 (9)

[a] Average values were calculated considering only the frames relevant to the proposed DPCs, *i.e.* those in which the following conditions are met simultaneously: (i) the distance between the **C**:H12a and $W1_C$:O (deprotonating base) atoms is ≤ 3.0 Å, and (ii) the distance in the hydrogen bonds that may facilitate the deprotonation of cation **C** via multiple proton transfer reactions is ≤ 2.0 Å. Standard deviations are given in parentheses. See Figure 6 of the main text for conventions on the atom labels.

[b] The entries under DPC(lifetime) are the percentages of the simulation time, during which the conditions for DPCs are met (according to the geometrical criteria listed above). They are given in parentheses in the order MD1-MD2-MD3. MD1-MD3 correspond to MD simulations with different initial velocity distributions.

[c] Not identified.

* Note: one may use different cutoff values for the relevant interatomic distances, for which it is still reasonable to expect that a multiple proton transfer reaction can occur (e.g. 2.5 Å instead of 2.0 Å for the $W1_C$:H-$W2_C$:O hydrogen bond). This leads to different numerical values for the average distances and DPC lifetimes; however, the conclusions on the deprotonation of cation **C** (as described in the main text) remain the same.

9

4. Cation F

Table S6

Average RMSD (with respect to the initial structure) for all carbon atoms of cation **F** in the MD simulations of the **TXS·F** complex[a]

MD	RMSD (Å)
MD1	1.2 (0.2)
MD2	0.8 (0.2)
MD3	1.1 (0.2)

[a] Standard deviations are given in parentheses. MD1-MD3 correspond to MD simulations with different initial velocity distributions.
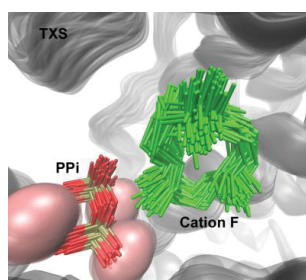
Figure S6. Superimposed structures of the **TXS·F** complex at intervals of 10 ps during MD1. MD2 and MD3 show similar superimposition profiles.
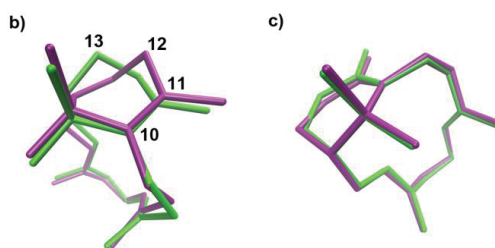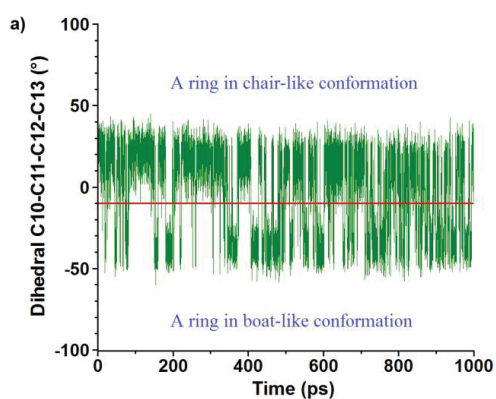
Figure S7. (a) Time evolution of the C10-C11-C12-C13 dihedral angle of cation **F** during the MD1 simulation of the **TXS·F** complex showing the flipping of the A ring between a chair-like (**F**$_{chair}$, dihedral > -10°) and boat-like (**F**$_{boat}$, dihedral < -10°) conformation. MD2 and MD3 have similar profiles. (b) **F**$_{chair}$ (green) and **F**$_{boat}$ (purple) superimposed. (c) HT-QM structure of cation **F** (purple) [1] superimposed with **F**$_{chair}$ (green, taken from the MD simulations of the **TXS·F** complex).

10

Table S7

Average values of the relevant interatomic distances and angles involved in the possible deprotonation paths of cation **F** (DPF) identified in the MD simulations of the **TXS·F** complex[a]

DPF (lifetime[b], %)	Distance ; Angle Definitions	Distance (Å) ; Angle (°) Values		
		MD1	MD2	MD3
DPF1$_{C20}$ (70 – 48 – 30)	**F**:H20-PPi:O1 ; **F**:(C20-H20)-PPi:O1	2.6 (0.2) ; 153 (13)	2.7 (0.2) ; 145 (15)	2.7 (0.2) ; 140 (12)
DPF2$_{C20}$ (27 – 9 – 3)	**F**:H20-W1$_F$:O ; **F**:(C20-H20)-W1$_F$:O	2.8 (0.2) ; 135 (15)	2.7 (0.2) ; 139 (16)	2.8 (0.2) ; 126 (11)
	W1$_F$:H-PPi:O1 ; W1$_F$:(O-H)-PPi:O1	1.7 (0.1) ; 165 (8)	1.7 (0.1) ; 165 (8)	1.7 (0.1) ; 166 (7)
DPF1$_{C2}$ (1 – 5 – 6)	**F**:H2-PPi:O1 ; **F**:(C2-H2)-PPi:O1	2.8 (0.2) ; 124 (12)	2.7 (0.2) ; 124 (14)	2.8 (0.2) ; 112 (11)
DPF2$_{C2}$ (71 – 61 – 84)	**F**:H2-W2$_F$:O ; **F**:(C2-H2)-W2$_F$:O	2.7 (0.2) ; 118 (13)	2.7 (0.2) ; 129 (12)	2.6 (0.2) ; 129 (13)
	W2$_F$:H-PPi:O1 ; W2$_F$:(O-H)-PPi:O1	1.7 (0.1) ; 166 (7)	1.7 (0.1) ; 165 (8)	1.7 (0.1) ; 166 (8)

[a] Average values were calculated considering only the frames in which the distance between the deprotonating base (PPi:O1, W1$_F$:O or W2$_F$:O) and the hydrogen atom to be abstracted from cation **F** (H20 or H2) is ≤ 3.0 Å. Regarding DPF2$_{C20/C2}$, the frames were selected by also requiring a distance ≤ 2.0 Å for the water-PPi hydrogen bond that may facilitate deprotonation of **F** via a double proton transfer reaction. Standard deviations are given in parentheses. See Figure 7 in the main text for conventions on the atom labels.

[b] The entries under DPF(lifetime) are the percentages of the simulation time, during which the conditions for DPFs are met (according to the geometrical criteria listed above). They are given in parentheses. They correspond to MD simulations with different initial velocity distributions.

* Note: one may use different cutoff values for the relevant interatomic distances, for which it is still reasonable to expect that the reactions in question can occur (e.g. 3.3 Å instead of 3.0 Å for **F**:H20-PPi:O1 and **F**:H20-W1$_F$:O). This leads to different numerical values for the average distances and DPC lifetimes; however, the conclusions on the deprotonation of cation **F** (as described in the main text) remain the same.

11

5. Cation D1

Table S8
Average RMSD (with respect to the initial structure) for all carbon atoms of cation **D1** in the MD simulations of the **TXS·D1** complex[a]

MD	RMSD (Å)
MD1	1.5 (0.4)
MD2	0.7 (0.2)
MD3	0.8 (0.2)

[a] Standard deviations are given in parentheses. MD1-MD3 correspond to MD simulations with different initial velocity distributions. The higher occurrence of $D1_{boat}$ during MD1 contributes to the higher RMSD value obtained for this simulation (see Table 1 of the main text).
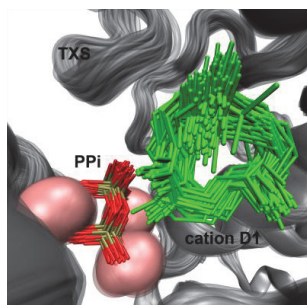
Figure S8. Superimposed structures of the **TXS·D1** complex at intervals of 10 ps during MD1. MD2 and MD3 show similar superimposition profiles.

12
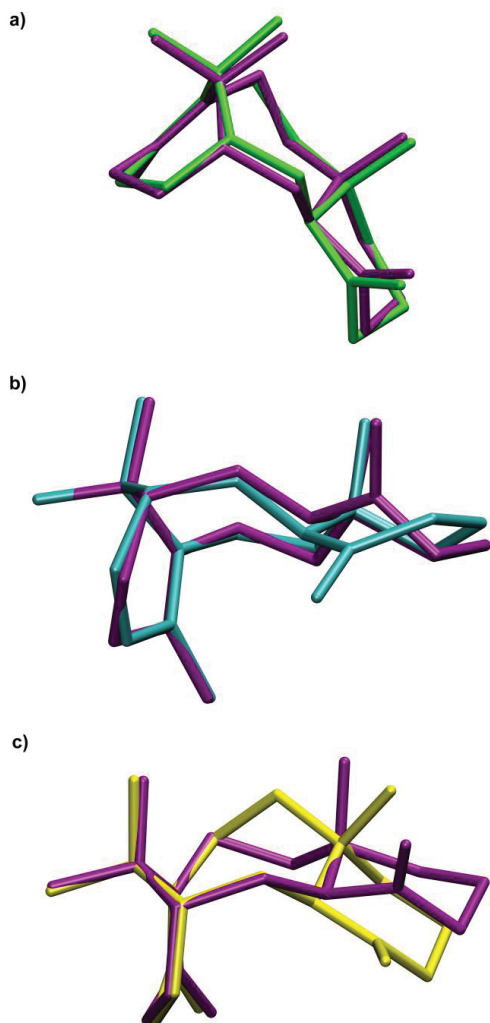
6. Cation E

Figure S9. E_{chair} (purple) superimposed with: (a) the HT-QM structure of cation **E** (green) [1], (b) $E_{t\text{-}boat}$ (cyan), and (c) E_{B1} (yellow).
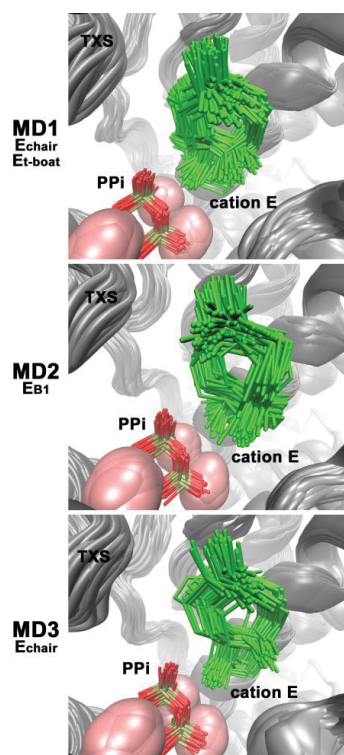
Figure S10. Superimposed structures of the **TXS·E** complex at intervals of 10 ps during MD1-MD3. The conformers of cation **E** identified in each simulation are indicated on the left.

13

7. References

1. Hong YJ, Tantillo DJ (2011) The Taxadiene-Forming Carbocation Cascade. J Am Chem Soc 133:18249–18256.

14

B. QM/MM Study of Taxadiene Synthase Catalysis

Jeaphianne P. M. van Rijn, Andrés M. Escorcia and Walter Thiel

I carried out all calculations and performed the subsequent analysis. I wrote the manuscript and created all included figures.

Journal of
COMPUTATIONAL
CHEMISTRY

FULL PAPER

QM/MM Study of the Taxadiene Synthase Mechanism

Jeaphianne P. M. van Rijn, Andrés M. Escorcia, and Walter Thiel ⓘ*

Combined quantum mechanics/molecular mechanics (QM/MM) calculations were used to investigate the reaction mechanism of taxadiene synthase (TXS). TXS catalyzes the cyclization of geranylgeranyl diphosphate (GGPP) to taxadiene (T) and four minor cyclic products. All these products originate from the deprotonation of carbocation intermediates. The reaction profiles for the conversion of GGPP to T as well as to minor products were calculated for different configurations of relevant TXS carbocation complexes. The QM region was treated at the M06-2X/TZVP level, while the CHARMM27 force field was used to describe the MM region. The QM/MM calculations suggest a reaction pathway for the conversion of GGPP to T, which slightly differs from previous proposals regarding the number of

reaction steps and the conformation of the carbocations. The QM/MM results also indicate that the formation of minor products via water-assisted deprotonation of the carbocations is highly exothermic, by about −7 to −23 kcal/mol. Curiously, however, the computed barriers and reaction energies indicate that the formation of some of the minor products is more facile than the formation of T. Thus, the present QM/MM calculations provide detailed insights into possible reaction pathways and into the origin of the promiscuity of TXS, but they do not reproduce the product distribution observed experimentally. © 2019 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.25846

Introduction

Taxol, commercially known as Paclitaxel, is a potent anticancer drug, which is widely used in the treatment of breast, lung, and ovarian cancer.[1,2] The delineation of the biosynthetic mechanism of taxol is an active field of research, and many steps of taxol biosynthesis from the acyclic universal diterpene precursor geranylgeranyl diphosphate (GGPP) have been identified.[3–5]

The first committed step on this pathway is the cyclization of GGPP to taxa-4(5),11(12)-diene (taxadiene, T), which is catalyzed by the enzyme taxadiene synthase (TXS, *Taxus brevifolia*) (see Fig. 1).[3,9] Being a class I terpene synthase, TXS initiates the catalysis using a trinuclear cluster of Mg^{2+} ions to bind GGPP and to trigger departure (ionization) of the pyrophosphate group (PPi), leading to the formation of a highly reactive carbocation, which subsequently undergoes rearrangement and cyclization reactions to form the product.[10,11] A wide range of varying product distributions from the promiscuous TXS is reported in the literature, showing several minor products like taxa-4(20),11(12)-diene (T1), verticillia-3(4),7(8),12(13)-triene (V), verticillia-4(20),7(8),11(12)-triene (V1) and verticillia-3(4),7(8), 11(12)-triene (V2) (see Fig. 1).[7,12–14] Although experimental trapping of GGPP-derived carbocation intermediates has not been possible, labeling experiments indicate the existence of cationic intermediates B, C, D, and E (Fig. 1).[12,13,15,16] Quantum mechanical (QM) studies of the hydrocarbon chain of GGPP[6,17] complemented these experimental findings leading to the proposed reaction mechanism of taxadiene formation shown in Figure 1 (referred to as the HT-QM mechanism, HT being a label for its authors).

Upon binding of the substrate and catalytic metal ions, the enzyme TXS is expected to undergo a conformational change from an open to a closed conformation.[10,18–22] However, in the available crystal structure of TXS (PDB ID 3P5R), the active

site is not fully closed.[10] Considering the closed structure of other terpene synthases, it is expected that in the catalytically active closed conformation the N terminal random coil segment (hereafter referred to as the NTRC, D80–L108) caps the active site, while the A–C (G570–H579), J–K (F837–E846), and H–α1 (Q770–S773) loops and the C-terminal portion of helix H (K760–G769) flank the mouth of the active site.[7,10,23,24] Unfortunately, crystal structures of closed TXS complexed with GGPP or carbocation analogues have not been reported. However, by applying homology modeling, two models of TXS in the closed conformation have been constructed up to now, which are referred to here as the FHM[24] and SHM[7] (i.e. the Freud and Schrepfer Homology Models).

For the construction of the FHM, sequence alignment, structure superposition, homology modeling, and loop modeling were employed using the closed form of bornyl diphosphate synthase from *Salvia officinalis* (BPPS, PDB ID 1N23[18]) as a template, while also considering the structures of 5-epi-aristolochene synthase from *Nicotiana tabacum* (PDB code 3M02[22]) and limonene synthase from *Mentha spicata* (PDB code 2ONG[25]).[24] The FHM was used to perform hybrid QM/MM and free energy simulations.[24,26] These suggest a six-step pathway for the biosynthesis of taxadiene from GGPP, where, in contrast to the HT-QM mechanism (Fig. 1), the direct pathway from C to D1 is (slightly) preferred over the indirect path via cation F.[24,26] The reaction was found to be strongly exothermic and downhill from cation A

[a] *J. P. M. van Rijn, Andrés M. Escorcia, W. Thiel*
Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1,
45470 Mülheim, Germany
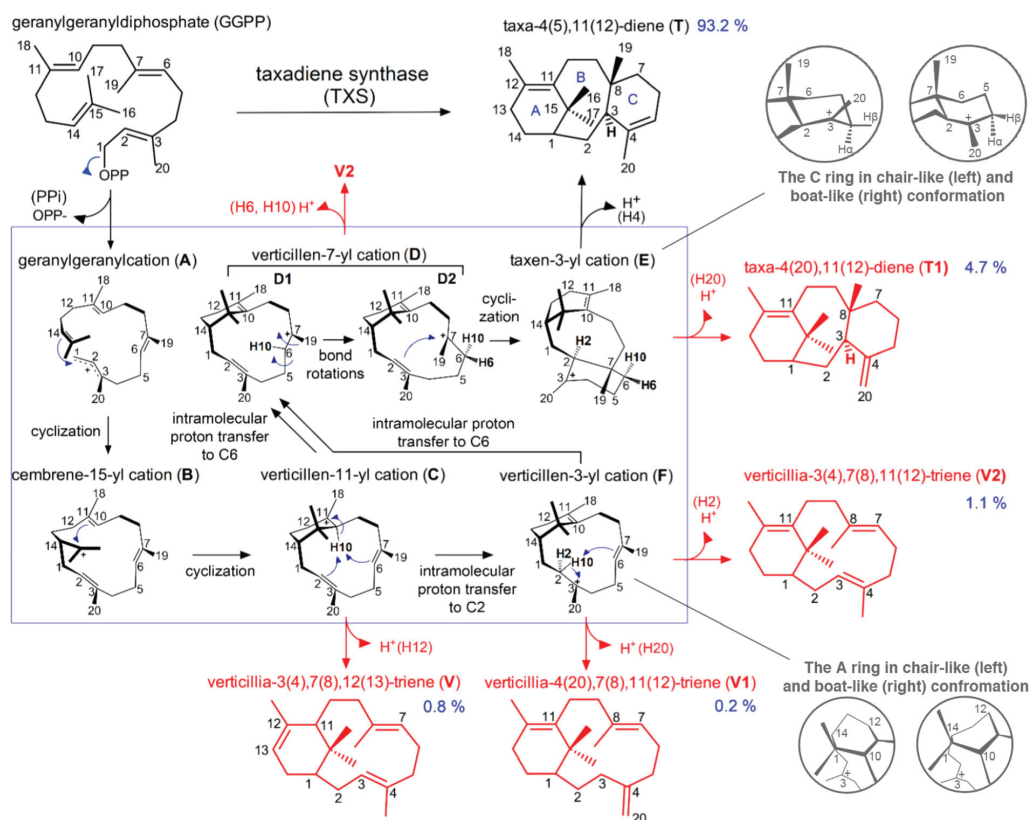E-mail: thiel@kofo.mpg.de

Figure 1. In black: the reaction mechanism of the TXS-catalyzed cyclization of GGPP to taxadiene from QM model calculations (HT-QM mechanism).[6] In red: minor products of TXS catalysis. In blue: the WT product yields as reported by Schrepfer et al.[7] and the taxadiene ring systems labels (A-C). In gray: illustrations of the chair- and boat-like conformation of the A ring (as observed for cations F, D) and the C ring (for cation E).[8] For the carbocation intermediates the numbering of GGPP is used, while minor products are numbered according to taxadiene convention. This figure is an adapted version of Figure 1 in reference [8]. [Color figure can be viewed at wileyonlinelibrary.com]

to taxadiene. A similar reaction profile has been reported for other terpene synthases.[19,27] A global free energy barrier of 24 kcal/mol was obtained, in good agreement with the experimentally determined k_{cat} of 0.0106 s^{-1} ($\Delta G^{\ddagger} \approx 20.1$ kcal/mol).[26] The reaction energetics for formation of side products was not addressed, although a qualitative explanation for the promiscuity of TXS was provided.[24,26] In the FHM active-site architecture, residue W753 was identified to be critical for stabilizing the carbocations through π-cation interactions.[26] The TXS W753H mutant exclusively produces cembrene A[7], which could be rationalized based on the FHM, further supporting the quality of this model.

To obtain the SHM, closed BPPS (PDB ID 1N20)[18] was used as a template for homology modeling followed by an energy minimization and a 10-ns MD simulation of the structure in water.[7,10] Models of the closed TXS complex, containing GGPP as well as all the carbocations in the HT-QM mechanism (hereafter denoted as ^{TXS}cation complexes), were created using an integrated approach of docking and molecular mechanics (MM) optimizations.[7] Each of these ^{TXS}cation complexes can

be distinctly derived from its precursor in the reaction cascade (Fig. 1), indicating that the HT-QM mechanism is feasible in the TXS environment.[7] The product distribution for the wild type (WT) enzyme was also reported and rationalized.[7] Based on the SHM, mutations were proposed and the predictive power of the SHM was confirmed by experimental site-directed mutagenesis data.[7]

Though BPPS was used as a template for both the FHM and the SHM, the highly flexible A–C (G570–S579) and J–K (F837–E846) loops[8] are positioned differently leading to a different active-site architecture in the two models.[7,24]

Previously, we performed MD simulations of enzyme-substrate complexes between the SHM of TXS and cations C, F, D1, and E to study their the dynamic behavior.[8] The MD simulations of the ^{TXS}D1 and ^{TXS}F complexes reveal a boat-like conformation of the A ring of these cations (cation$_{boat}$), in addition to the chair-like conformation (cation$_{chair}$) observed before in the gas phase[6,8] (see Fig. 1). In the ^{TXS}E complex, the A ring adopts a boat-like conformation.[6,12] Labeling experiments using deuterated GGPP indicate a twist-boat conformation of the C ring in cation E and

exclusive deprotonation of the β face of E:C4 to T.[12,13] In agreement with those findings, the MD simulations indicate that the formation of T in the TXS environment requires a conformational change of the C ring from a chair-like (^{TXS}E_C_chair) to a boat-like (^{TXS}E_C_boat) conformation (see Figs. 1 and 5b) (here "C_" indicates the conformation of the C ring, deviating from the convention used in ref [8]). In agreement with previous studies on terpene synthases we find PPi, which is considered to remain in the active site during the entire cyclization cascade, to play a key role in taxadiene synthase catalysis, through significant electrostatic interactions with the carbocation intermediates.[7,8,18,19,28–32] Judging from the average values of relevant interatomic distances, we proposed deprotonation paths for ^{TXS}F (deprotonation either directly by PPi or by a water-assisted proton transfer to PPi to produce either V1 or V2), for ^{TXS}C (water-assisted proton transfer to PPi yielding V), and for ^{TXS}E (direct deprotonation by PPi to T or T1) (see Fig. 1).[8]

In the present study, we go beyond MD simulations of the ^{TXS}cation complexes and compute complete energy profiles for the conversion of GGPP to T in the TXS environment using QM/MM calculations. Calculating the barriers of the proposed deprotonation pathways to the side products provides further insight into the promiscuity of TXS, and by investigating the conformational changes of the A and C rings we find a variant of the HT-QM mechanism.

Methods

QM calculations in the gas phase

To determine the effect of the enzyme environment on the computed QM/MM energy profiles of the GGPP to T conversion (see below), gas-phase QM calculations of the carbocation intermediates using density functional theory (DFT) serve as reference.[11,19,27,28] The gas-phase QM energies are denoted with a superscript GQ (^{GQ}cation). The DFT calculations were performed using the Gaussian09 program (for the reference, see the Supporting Information). Two sets of unconstrained geometry optimizations were carried out: (1) The reported HT-QM structures (Fig. 1)[6] were (re)optimized using different functionals (B3LYP with and without empirical dispersion corrections,[33,34] M06-2X,[35] and ωB97XD[36]) and basis sets (6–31+G(d,p)[37] and TZVP[38]) to quantify the effect of different basis sets and functionals (see Supporting Information Table S2). (2) The structures of cations C-E and the reaction products, as found in the QM/MM calculations, were optimized at the M06-2X/6–31+G(d,p) level (see Supporting Information Tables S3 and S4). Additionally, single-point (SP) calculations on the previously reported HT-QM structures[6] were performed with M06-2X/6–31+G(d,p) and at the DLPNO-CCSD(T)/def2-TZVPP[39,40] coupled cluster level (the latter using ORCA[41]) (see Supporting Information Table S2). Furthermore, CCSD(T) single-point calculations were done on the M06-2X/6–31+G(d,p) optimized structures of set 2 (see Supporting Information Tables S3 and S4). The reported energies include zero-point energy corrections, unless noted otherwise.

The reoptimized data show that energy differences between methods are small and mainly due to structural changes during optimization: structural differences compared to the HT-QM structures[6] are most evident in reoptimized cation A and decrease for cations appearing later in the cascade (see Fig. 1 and Supporting Information Table S2). Since the M06-2X functional has been shown to be suitable for computing carbocation reactions[19,24,26,28] it was chosen for use in the subsequent QM/MM calculations. The QM energies of the M06-2X single-point calculations are used for comparison with the catalytic pathway in the enzyme environment (Fig. 4). For more details on the QM results please see Section 2 of the Supporting Information.

QM/MM calculations of reaction profiles

For the QM/MM calculations of reaction profiles, we employed a static approach in analogy to previously reported studies of our group.[42–45] Representative snapshots were taken from the MD simulations of the ^{TXS}C complex (our previous study)[8] that served as starting points for the QM/MM calculations. Starting from cation C assures an appropriate fold of GGPP for the reaction to proceed to T. For the selection of these snapshots we considered: (1) the occurrence (lifetime) of the particular configuration of the system in the MD simulations, (2) the water network around PPi, (3) the orientation of the cation, and (4) the conformation of the enzyme, in particular concerning the NTRC segment and the concomitant positioning of PPi in the active site (see Supporting Information Fig. S1). A total of five snapshots were selected, which are labeled ^{TXS}C:W1E1, ^{TXS}C:W1E2, ^{TXS}C:W2E1, ^{TXS}C:W2E2, and ^{TXS}C:W1E2C (see Supporting Information Fig. S2). The first four snapshots feature a long MD lifetime combined with two different water networks (indicated by the labels W1 and W2) and two enzyme conformations (E1 & E2). W1 and W2 correspond to one or two water molecules mediating a hydrogen bond interaction between PPi and the adjacent residue Y835. E1 is characterized by an NTRC orientation that positions PPi close to the cation (the average PPi:O1-^{TXS}C:C1 distance in W1E1 and W2E1 is 3.2 ± 0.1 Å) and the formation of an S-H···π interaction between the thiol group of C830 and W753. In E2, the latter interaction is absent and the PPi:O1-^{TXS}C:C1 distance is 4.2 ± 0.8 Å (average of W1E2, W2E2, and W2E2C). ^{TXS}C:W1E2C has a different orientation of cation C in the binding pocket, which appeared for a shorter time in the MD simulation. See Figure 2 and Supporting Information Figures S1 and S2. In all five snapshots, the conformation of cation C is similar, with an RMSD from the average structure of 0.9 Å.

Six additional snapshots were taken from the MD simulations of the ^{TXS}F complex[8] for computation of all pathways to the side products V1 and V2 discussed in the section Introduction. For comparison to the energy profile obtained when starting from ^{TXS}C, a snapshot was taken from the MD simulations of ^{TXS}E, which was propagated backward.

The present ^{TXS}cation models comprise about 24,000 atoms and include about 5800 TIP3P water molecules. They contain closed TXS (6133 atoms), the carbocation (53 atoms), the catalytic Mg^{2+} ions (3 atoms), the neutralizing Na$^+$ ions (10 atoms), and PPi (9 atoms). The TXS part is composed of the structural motifs directly involved in the catalysis: the C-terminal domain
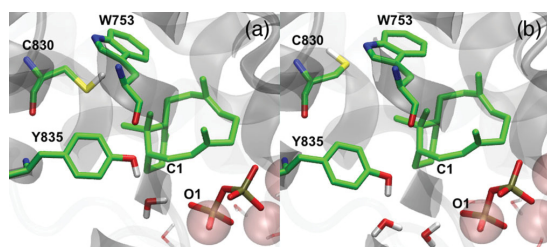
Figure 2. a) C:W1E1, with one water molecule between Y835 and PPi:O1; PPi:O1-^{TXS}C:C1 distance of 3.3 Å due to NTRC orientation E1; C830 interacting with W753, which plays a key role for the orientation of the C15 (CH$_3$)$_2$ moiety of GGPP. b) C:W2E2, with two water molecules between Y835 and PPi:O1; PPi:O1-^{TXS}C:C1 distance of 3.6 Å due to NTRC orientation E2; C830 points away from W753. In both structures most hydrogen atoms are omitted for clarity. See Supporting Information Figure S2 for setup C:W2E1 (E1 and two water molecules between Y835 and PPi:O1), setup C:W1E2 (E2 and one water molecule between Y835 and PPi:O1) and setup C:W1E2C (different orientation of C). [Color figure can be viewed at wileyonlinelibrary.com]

and part of the N-terminal domain (D80-L130 & N537-Q552) (see Supporting Information Fig. S1).[8,10] For a full account of the system setup, we refer the readers to our previous MD paper.[8]

The QM/MM calculations were done using the ChemShell program suite.[46,47] The QM part of the system was treated at the DFT level (M06-2X/TZVP) using the Gaussian09 software. Two different QM regions were used (Fig. 3). QM region I consists of the carbocation, the catalytic Mg^{2+} ions, and the PPi anion, with a total charge of +3 (QM1, 65 atoms). For calculations of deprotonation reactions involving water, QM region II was used, which consists of all QM1 components plus one water molecule (QM2, 68 atoms). The remainder of the system was treated at the MM level using the CHARMM27 force-field parameters.[48–50] The MM energies and gradients were computed by DL_POLY[51] as implemented in ChemShell. The interaction of the QM region with the point charges of the MM surrounding was handled by electrostatic embedding combined with the charge shift scheme.[52] Residues within a 6 Å radius of the initial positions of PPi, the carbocation, and the Mg^{2+} ions (active region) were unconstrained during the optimizations, while the rest of the system was kept fixed (Fig. 3).

After optimization of the ^{TXS}C snapshots, scans were carried out using a suitable reaction coordinate to obtain the pathway back to ^{TXS}GGPP and forward to ^{TXS}T and all side products (Fig. 1). For example, for the scans from ^{TXS}C to ^{TXS}F$_{chair}$ the difference between the C10 − H10 and H10 − C2 distances was chosen as reaction coordinate, while for the scans from ^{TXS}F$_{chair}$ to ^{TXS}F$_{boat}$ the dihedral angle C12-C13-C14-C15 was used (see Supporting Information Table S1 for the definition of all reaction coordinates). The highest point on a scan provided an initial guess for the corresponding transition state (TS) structure, which was subsequently optimized. In these TS optimizations a core region (of 7 to 53 atoms) was specified and treated using the P-RFO algorithm,[53,54] whereas the remaining non-frozen nuclei were treated by the L-BFGS algorithm.[55,56] All optimized structures were subjected to numerical force constant calculations in ChemShell to determine the vibrational modes and to characterize the optimized stationary points (one negative eigenvalue for
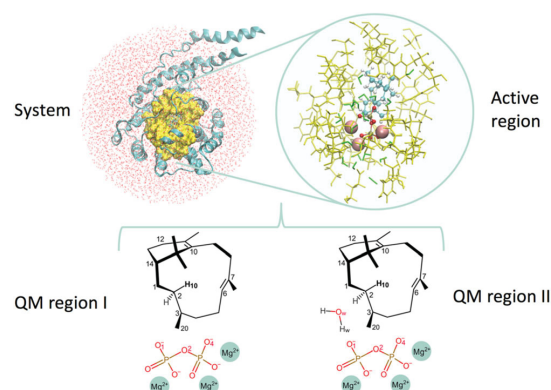
a TS, none for a minimum). Intrinsic reaction coordinate (IRC)-like calculations[57] and visual inspection ensured that a continuous pathway connected the computed stationary points.

Figure 3. Schematic representation of the QM/MM setup. [Color figure can be viewed at wileyonlinelibrary.com]

Results

GGPP → E QM/MM reaction profile

The GGPP → E QM/MM reaction profiles obtained with the different setups are similar for most reaction steps (see below). Therefore, we will present the results obtained with the representative setup C:W1E1 in detail and describe those for the other setups more briefly, focusing on relevant differences with respect to the C:W1E1 results.

Setup C:W1E1. A nine-step pathway was identified for the conversion of GGPP to T via C:W1E1, which is depicted in Figure 4.

The conformation of GGPP in the TXS environment (^{TXS}GGPP) is quite different from that predicted for the HT-QM mechanism in the gas phase (^{GQ}GGPP)[6] (see Fig. 5a). The C14 = C15 double bond is positioned closer to the C1-O11 bond in ^{TXS}GGPP (with C1···C14 and C1···C15 interatomic distances of 3.4 and 4.5 Å) than in ^{GQ}GGPP (C1···C14 distance of 6.5 Å[6]). Also, the C10···C15 interatomic distance is 1.5 Å shorter in ^{TXS}GGPP. Due to this preorganization in the enzyme environment, the hydrolysis of PPi and the anti-addition of C1 and C10 to the C14 = C15 double bond of ^{TXS}GGPP to give ^{TXS}C are a concerted process (see Figs. 1 and 4). With an energy barrier of 26.6 kcal/mol, the PPi loss and the subsequent cyclization step are rate-limiting, in line with findings for other terpene synthases as well as the FHM of TXS.[26,27]

The ^{TXS}C complex was found to rearrange more easily to ^{TXS}F than to ^{TXS}D1 in agreement with previous findings (Fig. 4).[6–8] This result contrasts recent studies that utilize the FHM of the TXS, where a slight preference for the direct proton transfer to C6 to form ^{TXS}D1 is found.[24,26] While the conversion of ^{GQ}C to ^{GQ}F is endothermic, it is facilitated thermodynamically in the enzyme, because the stabilizing effect of PPi through electrostatic interactions is larger in ^{TXS}F than in ^{TXS}C (see Supporting Information Fig. S3). Both ^{TXS}F$_{chair}$ and ^{TXS}F$_{boat}$ are stable complexes. The barrier for the conformational change of the A ring is 3.3 kcal/mol, with ^{TXS}F$_{boat}$ being 2.4 kcal/mol
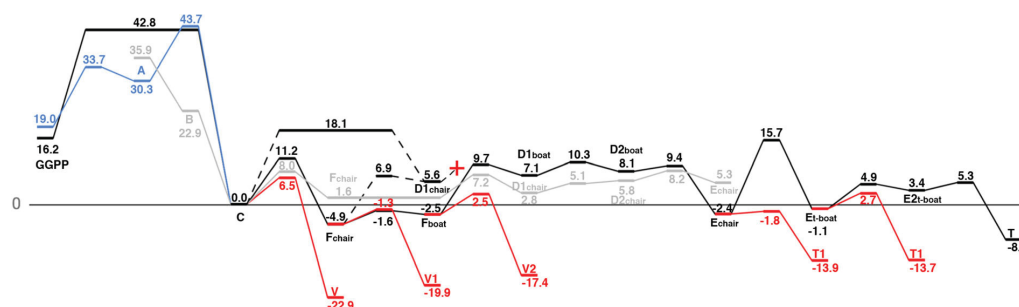
Figure 4. QM(M06-2X/TZVP)/MM(CHARMM) reaction profile of the TXS-catalyzed cyclization of GGPP to taxadiene of setup C:W1E1 (in black), GGPP-C for setup C:W2E2 (in blue) and minor products (red). All energy values (kcal/mol) are relative to the ^{TXS}C complex. In gray: gas-phase QM SP energies (M06-2X/6–31+G(d,p), relative to cation C) on HT-QM structures[6] of cation A-E. [Color figure can be viewed at wileyonlinelibrary.com]

less stable than ^{TXS}F$_{chair}$. This is consistent with the distribution observed in our previous MD simulations (^{TXS}F$_{chair}$ > 60% of the simulation time, versus ^{TXS}F$_{boat}$ < 40%).[8] ^{TXS}F$_{chair}$ is comparable to the HT-QM structure and apart from the A ring, so is ^{TXS}F$_{boat}$ (see Supporting Information Fig. S4).

Starting from cation F an intramolecular proton transfer of H10 from F:C2 to C6 leads to formation of cation D1. The H10-C6 distance is 2.3 Å in both ^{TXS}F$_{chair}$ and ^{TXS}F$_{boat}$ as well as in the HT-QM structure. Hence there is no clear preference for the conversion of ^{TXS}F to ^{TXS}D1 with the A ring in chair-like (barrier of 11.8 kcal/mol) or boat-like conformation (barrier of 12.2 kcal/mol). Because of the decrease in electrostatic interactions between PPi and the cationic center when going from ^{TXS}F to ^{TXS}D1 (see Supporting Information Figs. S3-S4 and Table S10), regardless of the conformation of the A ring, the F-D1 barrier is significantly higher in the enzyme than in the gas phase (5.6 kcal/mol[6]) and ^{TXS}D1 is about 10 kcal/mol higher in energy than ^{TXS}F. The ^{TXS}F$_{chair/boat}$→ ^{TXS}D1$_{chair/boat}$ forward reactions are favored over the ^{TXS}F$_{chair}$→^{TXS}C$_{chair}$ backward reaction, which has a barrier of 16.1 kcal/mol. ^{TXS}C$_{boat}$ was not observed in the MD simulations of the ^{TXS}C complex[8] and when optimized in the gas phase, ^{GQ}C$_{boat}$ converts to a chair-like A ring conformation (see Supporting Information Table S3). Because ^{TXS}C$_{boat}$ is not a minimum, the backward reaction from ^{TXS}F$_{boat}$ to ^{TXS}C$_{boat}$ goes to ^{TXS}C$_{chair}$ (see Supporting Information Fig. S5).

^{GQ}D1$_{boat}$ is stable in the gas phase, but 5.1 kcal/mol higher in energy than ^{GQ}D1$_{chair}$ (see Supporting Information Table S3). Regardless whether the A ring is in chair- or boat-like conformation, the complex of ^{TXS}D1 cannot form the C2-C7 σ-bond to generate the ^{TXS}E complex, since the C2···C7 interatomic distance is >4.0 Å.[6,8] To generate the productive conformer ^{TXS}D2 (see Fig. 1), the 12-membered ring of ^{TXS}D1 needs to undergo a conformational change along the C4-C5 and C5-C6 bonds to bring the positive charge on C7 sufficiently close to the C2 = C3 π-bond; the C2···C7 interatomic distance is 3.7 Å in ^{GQ}D2$_{chair}$.[6] In the enzyme we could find this conformational change for ^{TXS}D1$_{boat}$ but not for ^{TXS}D1$_{chair}$ despite using the same scan procedure. Apart from the A ring, both ^{TXS}D1$_{boat}$ and ^{TXS}D2$_{boat}$ are structurally similar to the HT-QM structures (see Supporting Information Fig. S4).

Conversion from ^{TXS}D2$_{boat}$ to ^{TXS}E$_{C\_chair}$ (A ring in boat- and C ring in chair-like conformation; see Fig. 1) has a low barrier

(1.3 kcal/mol) and a reaction energy of −10.5 kcal/mol. The conversion of ^{TXS}GGPP to ^{TXS}E$_{C\_chair}$ is exothermic by 18.6 kcal/mol. Structurally, ^{TXS}E$_{C\_chair}$ aligns well with the HT-QM structure (see Supporting Information Fig. S4). Assessment of the interactions on the basis of interatomic distances in previous computational work[7,8] suggested ^{TXS}E$_{C\_chair}$ to be the energetic minimum of the carbocation intermediates. However, ^{TXS}F$_{boat}$ and ^{TXS}F$_{chair}$ are more stable than the ^{TXS}E$_{C\_chair}$ complex, as the stabilizing effect of PPi on these complexes is comparable (see Supporting Information Fig. S3) while cation F is intrinsically more stable (see Fig. 4).[6]

In the MD simulations of the ^{TXS}E complex, the C ring was found in both chair (^{TXS}E$_{C\_chair}$) and twist-boat (^{TXS}E$_{C\_boat}$) conformation (see Fig. 5b).[8] For setup C:W1E1, the barrier from ^{TXS}E$_{C\_chair}$ to ^{TXS}E$_{C\_boat}$ is 18.1 kcal/mol, with the ^{TXS}E$_{C\_boat}$ complex being about 1 kcal/mol less stable. In the gas phase, cation ^{GQ}E is 4.7 kcal/mol more stable when the C ring is in a twist-boat conformation (see Supporting Information Table S3). Direct formation of ^{TXS}E$_{C\_boat}$ in the enzyme, by internal rotation around the C3-C4 bond in ^{TXS}D2$_{boat}$, would lead to the wrong orientation of H2 in the final product (see Supporting Information Fig. S6). A rotation of ^{TXS}E$_{C\_boat}$ in the binding pocket, with a barrier of 6.0 kcal/mol, leads to the ^{TXS}E2$_{C\_boat}$ complex that is 4.5 kcal/mol less stable (see Fig. 5b). Favorably, the PPi:O1- H4β distance in ^{TXS}E2$_{C\_boat}$ is 3.2 Å, which is shorter than the distances of 4.5 Å and 4.3 Å found in the ^{TXS}E$_{C\_chair}$ and ^{TXS}E$_{C\_boat}$ complexes, respectively (see Supporting Information Table S8).

Other setups. Figure 4 also shows the reaction profile of the ^{TXS}GGPP → ^{TXS}C transformation as obtained for setup W2E2. When the NTRC is in E2 conformation, ^{TXS}GGPP first converts to ^{TXS}A, which then rearranges to ^{TXS}C, whereas for E1 a concerted pathway (as discussed for snapshot W1E1) is found (Fig. 4, Supporting Information Table S6). For snapshots with NTRC conformation E1, the thiol group of C830 interacts with W753 and is pointing towards C15(CH$_3$)$_2$ (see Fig. 2 and Supporting Information Fig. S7). Therefore, the C1···C14 and C1···C15 distances are about 0.3 Å shorter for these snapshots, than for those with E2 where the C830 thiol does not point towards C15(CH$_3$)$_2$ (see Fig. 2 and Supporting Information Figs. S2, S7, and Table S5). Increasing the ^{TXS}GGPP:O1-C1 distance for formation of ^{TXS}A
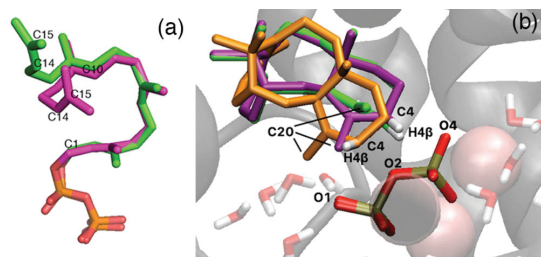
Figure 5. a) Overlay of HT-QM model of GGPP[6](green) and GGPP:W1E1 (magenta). (b) ^{TXS}E complexes identified in the QM/MM calculations: $^{TXS}E_{C\_chair}$ (cation in green) originates from the $^{TXS}D2_{boat}$ complex. $^{TXS}E_{C\_boat}$ (cation in purple) results from a conformational change of the C ring to boat-like conformation. $^{TXS}E2_{C\_boat}$ (cation in orange) originates from a slight rotation of the E_{C_boat} conformer. Hydrogen atoms are omitted for clarity. [Color figure can be viewed at wileyonlinelibrary.com]

brings C1 closer to C14. For E1 this leads directly to ^{TXS}C formation, while for E2 the larger C1···C14 distance leads to ^{TXS}A being a minimum.

The conformation of A in ^{TXS}A differs from that of cation ^{GQ}A,[6] with C15(CH$_3$)$_2$ being positioned closer to C10 by about 1.0 Å (C10-C15 distance) and rotated further than in the gas phase (see Supporting Information Fig. S4). Due to this preorganization, ^{TXS}B is not a minimum, which would explain the absence of cembrene A (CM) as a side product of the TXS catalysis (see Supporting Information Fig. S11).[7] Optimization from $^{TXS}TS(A-C)$ towards ^{TXS}C passes through a structure like cation ^{GQ}B (see Supporting Information Fig. S8). Also the transition states from both $^{TXS}GGPP$ and ^{TXS}A to ^{TXS}C resemble cation ^{GQ}B (see Supporting Information Fig. S4).

The $^{TXS}A \rightarrow {}^{TXS}C$ transformation has an average barrier of 11.1 ± 2.0 kcal/mol; the imaginary frequency corresponds to the formation of the C1-C14 σ-bond (Fig. 1 and Supporting Information Table S6). The formation of two new σ-bonds in cation C results in a drop in the relative QM/MM energy of about 30 kcal/mol (see Fig. 4 and Supporting Information Table S6), despite a significant reduction in the stabilizing effect of PPi (Supporting Information Fig. S3), due to the increased distance between the center of positive charge and PPi:O1 from 2.8 ± 0.2 Å in ^{TXS}A to 6.4 ± 0.3 Å in ^{TXS}C (for setups with E2).

From ^{TXS}C to $^{TXS}E_{C\_chair}$ the energy profiles and the stationary points of all the setups have an average standard deviation of about 3 kcal/mol (see Supporting Information Table S6), but the overall trends for all setups are comparable to those of the previously discussed setup C:W1E1.

The average barrier for $^{TXS}E_{C\_chair} \rightarrow {}^{TXS}E_{C\_boat}$ is 19.2 ± 1.5 kcal/mol for snapshots with E1, while snapshots with E2 have a barrier of 9.3 ± 1.7 kcal/mol. This is attributed to the steric hindrance between C20 and PPi being larger in the former than in the latter (see Supporting Information Tables S7, S8 and Fig. S10). A rotation of $^{TXS}E_{C\_boat}$ in the binding pocket, with barriers of 4–15 kcal/mol, leads to the $^{TXS}E2_{C\_boat}$ complex for all but the two W2E2 snapshots. For the two W2E2 snapshots scanning from $^{TXS}E_{C\_chair}$ to $^{TXS}E_{C\_boat}$ not only affects the C ring, but results in a rotation of the entire cation (see Supporting Information Fig. S10). When rotating the W2E2 $^{TXS}E_{C\_boat}$ complexes to $^{TXS}E2_{C\_boat}$, they convert back to $^{TXS}E_{C\_chair}$.

From the MD simulations of ^{TXS}E, an additional snapshot of $^{TXS}E2_{C\_boat}$ was taken, W1E2E (Supporting Information Fig. S10). Propagating the reaction backward from $^{TXS}E2_{C\_boat}$ to $^{TXS}GGPP$ yields an energy profile that is more in line with the FHM results[24,26] (see Supporting Information Table S9).

QM/MM reaction energies of the formation of the (side) products

We analyzed the expected preferred deprotonation pathways of the carbocations as discussed in the Introduction (see Supporting Information Table S1).[8] The results for W1E1 are shown in Figure 4. Supporting Information Tables S11 and S12 provide the average barriers and reaction energies for deprotonation and describe how the averages are calculated, for all setups.

In $^{TXS}E_{C\_chair}$ and $^{TXS}E_{C\_boat}$ deprotonation of the α face of E:C4 by PPi:O4 (see Fig. 5b) to yield T is feasible and favored over deprotonation of the β face by PPi:O4 or PPi:O1 (Supporting Information Table S11). After rotation to $^{TXS}E2_{C\_boat}$, deprotonation on the β face (as found in experiments[12,13]) by PPi:O1 becomes favorable with a barrier of 2.7 ± 1.7 kcal/mol (Supporting Information Table S11). Deprotonation of $^{TXS}E_{C\_chair}$:C20 and $^{TXS}E_{C\_boat}$:C20 by PPi:O1 to yield T1 has barriers of 2–17 kcal/mol (Supporting Information Table S11). Still, the energy barrier for T formation from the $^{TXS}E2_{C\_boat}$ complex is similar to or lower than that of T1 formation from the other ^{TXS}E complexes. Also for setup W1E2E (Supporting Information Fig. S10) the barrier to ^{TXS}T is 3.5 kcal/mol lower than the barrier to $^{TXS}T1$ (see Supporting Information Table S9). Gas-phase single-point coupled cluster calculations (CCSD(T)) of the different (side) products indicate that $^{GQ}T_{C\_boat}$ is the thermodynamically most stable product (see Supporting Information Table S4).

Both $^{TXS}F_{chair}$ and $^{TXS}F_{boat}$ can be converted to $^{TXS}V1$ or $^{TXS}V2$, with similar barriers from both conformations (see Supporting Information Table S12); they are therefore not discussed separately. V1 formation through direct deprotonation of ^{TXS}F:C20 by PPi:O1 has an average barrier of 4.4 ± 1.2 kcal/mol, and the process is exothermic by -12.2 ± 2.8 kcal/mol. For V2 formation the direct deprotonation of ^{TXS}F:C2 by PPi:O1 (one snapshot) is barrierless, while the water-assisted deprotonation of ^{TXS}F:C2 by PPi:O1 has a barrier of 2.7 ± 2.6 kcal/mol. V2 formation is exothermic by -18.2 ± 2.0 kcal/mol.

The water-assisted deprotonation of ^{TXS}C:C12 by PPi:O1 to form ^{TXS}V was found to have a low barrier of 5.6 ± 3.0 kcal/mol and to be exothermic by -23.4 ± 2.4 kcal/mol.

For W1E2E:C and W1E2E:F both the direct and water-mediated deprotonation by PPi:O1 to the side products seem unlikely (based on distances, see Supporting Information Table S13).

Discussion

Overall, our QM/MM calculations explain the promiscuity of the TXS but not the product distribution. The barriers to the side products are often lower than those to the next reaction step of the carbocation cascade (Fig. 4 and Supporting Information Tables S6, S11-S12). For example, the barrier to ^{TXS}V from ^{TXS}C

is over 5 kcal/mol lower than the barrier to $^{TXS}F_{chair}$. Also the barriers from ^{TXS}F to $^{TXS}V1$ and $^{TXS}V2$ are over 9 kcal/mol lower than those for conversion to $^{TXS}D1$. Formation of ^{TXS}T only becomes favorable over $^{TXS}T1$ formation after multiple conformational changes of ^{TXS}E with substantial barriers (6–20 kcal/mol), and not at all for the two W2E2 snapshots (Supporting Information Tables S7 and S11). Meanwhile, deprotonation of $^{TXS}E_{C\_chair}$ to $^{TXS}T1$ is found to be feasible for all snapshots, having barriers of 8.6 ± 5.9 kcal/mol (Supporting Information Table S11). Moreover, complexes of ^{TXS}E are higher in energy than complexes of ^{TXS}F, which would suggest that products arising from ^{TXS}F should dominate over products T and T1 (Fig. 4 and Supporting Information Table S6).

A possible explanation for the perceived preference for side product formation in our calculations could be that our model contains too many water molecules. The active-site water molecules in TXS influence the orientation of the cations and enable the low-barrier water-assisted deprotonation paths that can prematurely terminate the reaction. To our knowledge there is no experimental evidence on the number of water molecules remaining in the binding pocket of the enzyme after closing, though the involvement of water as a base in the catalysis of terpene synthase reactions (except for enzymes generating hydroxylated products) is an uncommon notion.[58] However, the large active-site volume of TXS could mean that TXS is different from other terpene synthases.[8,10] An active-site water has been previously identified to mediate the deprotonation of the camphyl carbocation to produce camphene during BPPS catalysis.[19]

Considering dynamical effects, the probability for premature deprotonation of the carbocations might be lower than that of the cations progressing to the next reaction step of the cascade. In the MD simulations, the deprotonation paths for ^{TXS}C and ^{TXS}F do not occur 100% of the time, while the required geometrical criteria for the cations to progress to the next reaction step of the carbocation cascade are met all along the trajectories.[8] So, although the barriers and reaction energies for formation of the side products are favorable, there might be a higher number of TXS cation complexes leading to formation of taxadiene than to formation of the side products, due to the dynamics.[59,60] Moreover, it has been shown for terpene synthases that though ionization of the diphosphate ester bond is the rate-limiting chemical step, product release is the rate-limiting step in the overall reaction.[27,61,62] Considering that the taxadiene complex is 1–15 kcal/mol less stable then the other product complexes (Supporting Information Tables S6-S7, S11-S12), taxadiene is expected to be released more easily, which may contribute to the higher yields of T observed experimentally.

Our QM/MM results on TXS catalysis (Fig. 4 and Supporting Information Table S6) deviate from recent QM/MM results obtained with the FHM.[26] Our pathway is less downhill from ^{TXS}A, shows a preference for the two-step pathway to ^{TXS}D via cation F, does not include ^{TXS}B, and does include conformational changes of the A and C rings (Fig. 4). The presence of ^{TXS}F in the pathway accounts for side products V1 and V2, while the absence of ^{TXS}B is in line with CM not being a side product of the TXS catalysis. The differences in the results from the present and previous[26] QM/MM studies may partly be due to the

different methods employed for computing the reaction profiles (static QM/MM vs. free energy QM/MM calculations). We believe, however, that the differences between the underlying structural models are more important: as outlined in the Introduction, the FHM and SHM were built using different setup procedures.[7,8,24] The published material on the FHM indicates that this model differs from the SHM in having a different number of active-site water molecules close to the carbocations (one vs. four), a different orientation of the substrate in the binding pocket (e.g., with respect to residue Y841), and a different structure around the active site (e.g., with respect to the positioning of the A–C (G570–H579) and J–K (F837–E846) loops); see Section 5 of the Supporting Information for more details.

The currently computed QM/MM reaction profiles for the five chosen snapshots give an internally consistent qualitative scenario for TXS catalysis but they also differ appreciably in a quantitative sense (see section Results). Furthermore, the additional snapshot W1E2E taken from the MD simulations of ^{TXS}E (Supporting Information Fig. S10) yields an energy profile that differs rather strongly from the others and is more in line with the FHM results[26] (see Supporting Information Table S9); we note that the orientation of the cation with respect to PPi differs in the W1E1 and W1E2E snapshots (see Supporting Information Table S10). In general, it is very difficult to predict the correct bound state of the substrate, intermediates, and product in terpene synthases.[27,63,64] This is mainly due to the absence of hydrogen bonds between the ligand and the enzyme that could serve as anchors to keep the former in place.[64] This may be especially challenging for TXS considering its very large active site.[10] Thus, though the structures of the TXS·cation complexes used in this study were obtained from a commonly employed docking[7] and MD procedure[8], it could still be possible that relevant orientations of the cations have been missed. The sensitivity of our static QM/MM results with regard to the chosen snapshot suggests that it would seem worthwhile to perform QM/MM dynamics simulations with the SHM, which might be helpful to rationalize the product distribution of TXS.

Conclusions

In this study, QM/MM calculations have been used to investigate the reaction mechanism of the cyclization of GGPP to taxadiene (T) and four side products catalyzed by TXS. They are based on a previously constructed model (SHM) of the reactive closed conformation of TXS.[7]

The reaction pathway for the conversion of GGPP to T found in our calculations differs from previous proposals. It contains additional steps for conformational changes of the A ring in ^{TXS}F and the C ring in ^{TXS}E as well as a rotation of cation E. Additionally, depending on the conformation of the enzyme, the $^{TXS}A \rightarrow {}^{TXS}C$ or the $^{TXS}GGPP \rightarrow {}^{TXS}C$ conversions may be concerted. The suggestion that ^{TXS}B is not a minimum on the PES contrasts other computational findings,[6,26] but seems consistent with experimental evidence: stereochemical labeling experiments[12,13] indicated that the GGPP→C transformation could be a concerted process, and cembrene A is not observed in the product distribution.[7]

According to the QM/MM calculations, formation of minor products via water-assisted deprotonation of carbocation intermediates, as proposed in our MD study,[8] is energetically favorable, with barriers of 0–20 kcal/mol and reaction energies of −7 to −23 kcal/mol. We note, however, that water-assisted deprotonation is an uncommon notion for terpene synthases.[58] Our calculations identify routes to the observed side products and thus explain the promiscuity of the enzyme, but they do not reproduce the observed product distribution since the barriers to the side products are often lower than those to the next reaction step of the carbocation cascade. However, it is conceivable that dynamical effects might disfavor the side reactions and that taxadiene is released more easily than the other products (see section Discussion).

The two published models of TXS in the closed conformation, the SHM[7] and the more recent FHM,[24] differ appreciably and give rise to different energy profiles. Both models, however, might reflect reality. The QM/MM free energy study with the FHM[26] provides an intriguing overall mechanistic scenario for TXS-catalyzed taxadiene formation, while our static QM/MM calculations offer detailed insight into the promiscuity of TXS. In this sense, the two studies are complementary to each other. It may be possible that the FHM and SHM reflect states of the TXS catalysis with high and low population that favor formation of taxadiene and of the side products, respectively.

From a methodological point of view, the present work underlines the pronounced sensitivity of the TXS-catalyzed carbocation rearrangements to the enzyme environment. This is mainly due to strong electrostatic interactions that depend on the positioning, orientation, and conformation of the cation as well as the active-site architecture and the presence of active-site water molecules. The unusually large sensitivity of the computed reaction profiles for TXS catalysis exemplifies the known limitations of the static QM/MM approach to enzyme reactivity.

[1] R. N. Patel, *Annu. Rev. Microbiol.* **1998**, *52*, 361.
[2] G. M. Cragg, *Med. Res. Rev.* **1998**, *18*, 315.
[3] S. Malik, R. M. Cusidó, M. H. Mirjalili, E. Moyano, J. Palazón, M. Bonfill, *Process Biochem.* **2011**, *46*, 23.
[4] S. Howat, B. Park, I. S. Oh, Y.-W. Jin, E.-K. Lee, G. J. Loake, *N. Biotechnol.* **2014**, *31*, 242.
[5] S. C. Roberts, *Nat. Chem. Biol.* **2007**, *3*, 387.
[6] Y. J. Hong, D. J. Tantillo, *J. Am. Chem. Soc.* **2011**, *133*, 18249.
[7] P. Schrepfer, A. Buettner, C. Goerner, M. Hertel, J. van Rijn, F. Wallrapp, W. Eisenreich, V. Sieber, R. Kourist, T. Brück, *Proc. Natl. Acad. Sci.* **2016**, *113*, E958.
[8] A. M. Escorcia, J. P. M. van Rijn, G. J. Cheng, P. Schrepfer, T. B. Brück, W. Thiel, *J. Comput. Chem.* **2018**, *39*, 1215.
[9] A. E. Koepp, M. Hezari, J. Zajicek, B. S. Vogel, R. E. LaFever, N. G. Lewis, R. Croteau, *J. Biol. Chem.* **1995**, *270*, 8686.
[10] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson, *Nature* **2011**, *469*, 116.
[11] D. J. Tantillo, *Nat. Prod. Rep.* **2011**, *28*, 1035.
[12] Y. Jin, D. C. Williams, R. Croteau, R. M. Coates, *J. Am. Chem. Soc.* **2005**, *127*, 7834.
[13] Q. Jin, D. C. Williams, M. Hezari, R. Croteau, R. M. Coates, *J. Org. Chem.* **2005**, *70*, 4667.
[14] T. A. Pemberton, M. Chen, G. G. Harris, W. K. W. Chou, L. Duan, M. Köksal, A. S. Genshaft, D. E. Cane, D. W. Christianson, *Biochemistry* **2017**, *56*, 2010.
[15] S. Y. Chow, H. J. Williams, Q. Huang, S. Nanda, A. I. Scott, *J. Org. Chem.* **2005**, *70*, 9997.
[16] D. C. Williams, B. J. Carroll, Q. Jin, C. D. Rithner, S. R. Lenger, H. G. Floss, R. M. Coates, R. M. Williams, R. Croteau, *Chem. Biol.* **2000**, *7*, 969.
[17] P. Gutta, D. J. Tantillo, *Org. Lett.* **2007**, *9*, 1069.
[18] D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau, D. W. Christianson, *Proc. Natl. Acad. Sci.* **2002**, *99*, 15375.
[19] D. T. Major, M. Weitman, *J. Am. Chem. Soc.* **2012**, *134*, 19454.
[20] D. W. Christianson, *Curr. Opin. Chem. Biol.* **2008**, *12*, 141.
[21] M. W. van der Kamp, J. Sirirak, J. Žurek, R. K. Allemann, A. J. Mulholland, *Biochemistry* **2013**, *52*, 8094.
[22] C. M. Starks, K. Back, J. Chappell, J. P. Noel, *Science* **1997**, *277*, 1815.
[23] D. C. Williams, M. R. Wildung, A. Q. Jin, D. Dalal, J. S. Oliver, R. M. Coates, R. Croteau, *Arch. Biochem. Biophys.* **2000**, *379*, 137.
[24] Y. Freud, T. Ansbacher, D. T. Major, *ACS Catal.* **2017**, *7*, 7653.
[25] D. C. Hyatt, B. Youn, Y. Zhao, B. Santhamma, R. M. Coates, R. B. Croteau, C. Kang, *Proc. Natl. Acad. Sci.* **2007**, *104*, 5360.
[26] T. Ansbacher, Y. Freud, D. T. Major, *Biochemistry* **2018**, *57*, 3773.
[27] M. Dixit, M. Weitman, J. Gao, D. T. Major, *ACS Catal.* **2017**, *7*, 812.
[28] D. T. Major, *ACS Catal.* **2017**, *7*, 5461.
[29] M. Xu, P. R. Wilderman, R. J. Peters, *Proc. Natl. Acad. Sci.* **2007**, *104*, 7397.
[30] K. Zhou, R. J. Peters, *Chem. Commun.* **2011**, *47*, 4074.
[31] E. Y. Shishova, L. Di Costanzo, D. E. Cane, D. W. Christianson, *Biochemistry* **2007**, *46*, 1941.
[32] Y. J. Hong, D. J. Tantillo, *Org. Biomol. Chem.* **2010**, *8*, 4589.
[33] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[34] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.
[35] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.
[36] J. D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
[37] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
[38] A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829.
[39] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2016**, *144*, 024109.
[40] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
[41] F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73.
[42] S. Metz, W. Thiel, *J. Am. Chem. Soc.* **2009**, *131*, 14885.
[43] I. Polyak, M. T. Reetz, W. Thiel, *J. Am. Chem. Soc.* **2012**, *134*, 2732.
[44] B. Karasulu, M. Patil, W. Thiel, *J. Am. Chem. Soc.* **2013**, *135*, 13400.
[45] A. M. Escorcia, K. Sen, M. C. Daza, M. Doerr, W. Thiel, *ACS Catal.* **2017**, *7*, 115.
[46] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1.
[47] ChemShell, a Computational Chemistry Shell, see www.chemshell.org.
[48] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.

B. Paper II

[49] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wio rkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.

[50] A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400.

[51] W. Smith, T. R. Forester, *J. Mol. Graph.* **1996**, *14*, 136.

[52] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580.

[53] A. Banerjee, N. Adams, J. Simons, R. Shepard, *J. Phys. Chem.* **1985**, *89*, 52.

[54] J. Baker, *J. Comput. Chem.* **1986**, *7*, 385.

[55] J. Nocedal, *Math. Comput.* **1980**, *35*, 773.

[56] D. C. Liu, J. Nocedal, *Math. Prog.* **1989**, *45*, 503.

[57] Further details about this method in SI Section 1.

[58] D. W. Christianson, *Chem. Rev.* **2017**, *117*, 11570.

[59] K. Arora, C. L. B. Iii, In Dyn. Enzyme Catal; J. Klinman, S. H. Schiffer, Eds., Springer, Berlin, Heidelberg, **2013**, p. 165.

[60] G. G. Hammes, S. J. Benkovic, S. Hammes-Schiffer, *Biochemistry* **2011**, *50*, 10422.

[61] D. E. Cane, H. T. Chiu, P. Liang, K. S. Anderson, *Biochemistry* **1997**, *36*, 8332.

[62] J. R. Mathis, K. Back, C. Starks, J. Noel, C. D. Poulter, J. Chappell, *Biochemistry* **1997**, *36*, 8340.

[63] M. Dixit, M. Weitman, J. Gao, D. T. Major, *ACS Catal.* **2018**, *8*, 1371.

[64] D. T. Major, Y. Freud, M. Weitman, *Curr. Opin. Chem. Biol.* **2014**, *21*, 25.

106

Supporting Information

QM/MM study of the taxadiene synthase mechanism

Jeaphianne P. M. van Rijn, Andrés M. Escorcia and Walter Thiel*

[*]W.T.: Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim, Germany. Email: thiel@kofo.mpg.de

Content:

1. Method and system setup

Table S1 Reaction coordinates used in the QM/MM PES scans to obtain the reaction profiles of the conversion of ^{TXS}GGPP to ^{TXS}T and side products[1]

Reaction step	Reaction coordinate used for scan
GGPP-A	O1-C1 distance
A-C	C1-C14 distance
C-F$_{chair}$	(C10-H10) – (H10-C2) distance difference
C-D1$_{chair}$	(C10-H10) – (H10-C6) distance difference
F$_{chair}$-F$_{boat}$	Dihedral angle C12-C13-C14-C15
F$_{chair}$-D1$_{chair}$ / F$_{boat}$-D1$_{boat}$	(C2-H10) – (H10-C6) distance difference
D1$_{boat}$-D2$_{boat}$	Dihedral angles C3-C4-C5-C6 and C4-C5-C6-C7 (simultaneously)
D2$_{boat}$-E$_{C\_chair}$	C2-C7 distance
E$_{C\_chair}$-E$_{C\_boat}$	Dihedral angle C20-C3-C4-C5
E$_{C\_boat}$-E2$_{C\_boat}$	O1-C4:Hβ distance
E2$_{C\_boat}$ -T	O1- C4:Hβ distance
E2$_{C\_boat}$ -T1	(C20:H^2-C20) - (O1-C20:H^2) distance difference
C-H$_2$O-V	(C12:H^2-C12) - (O$_w$-C12:H^2) distance difference [3]
F$_{chair}$/F$_{boat}$-V1	(C20:H^2-C20) - (O1-C20:H^2) distance difference
F$_{boat}$ -V2	(C2:H^2-C2) - (O1-C2:H^2) distance difference
F$_{chair}$/F$_{boat}$-H$_2$O-V2	(C2:H^2-C2) - (O$_w$ -C2:H^2) distance difference [3]

[1] Atom labels can be found in Figures 1 and 3 of the main text
[2] Hydrogen atom closest to base (PPi:O1).
[3] Water-mediated proton transfer reaction from cation to PPi. The concomitant decrease of the H$_w$-PPi:O1 distance occurred spontaneously when varying the indicated reaction coordinate.

IRC-like calculation

An IRC-like calculation makes use of an approximate IRC procedure. In an IRC-like calculation, a fraction of the normal mode eigenvector corresponding to the imaginary frequency of the transition state is added to (or subtracted from) the structure of the transition state. The new structure is subjected to an unconstrained geometry optimization, and the resulting structure is visually inspected to confirm that it is the reactant or product.

2
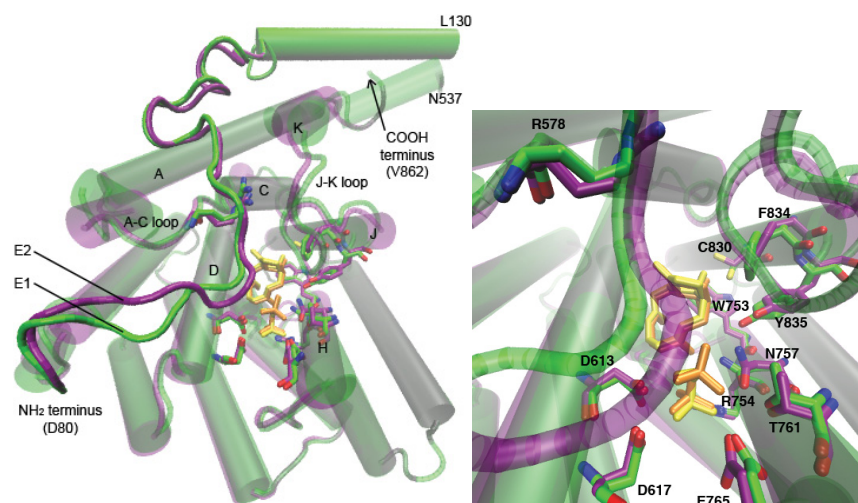
Figure S1. Left: Structure of the entire enzyme for snapshot W1E1:**C** (green) and W2E2:**C** (purple) comprising the C-terminal domain (S553-V862) and part of the N-terminal domain (D80-L130 and N537-Q552). Right: Zoom of the active site with a selection of the residues responsible for positioning of the PPi anion (D613, D617, R754, N757, T761, E765) or positioning of cation **C** (W753, C830, F834, Y835).
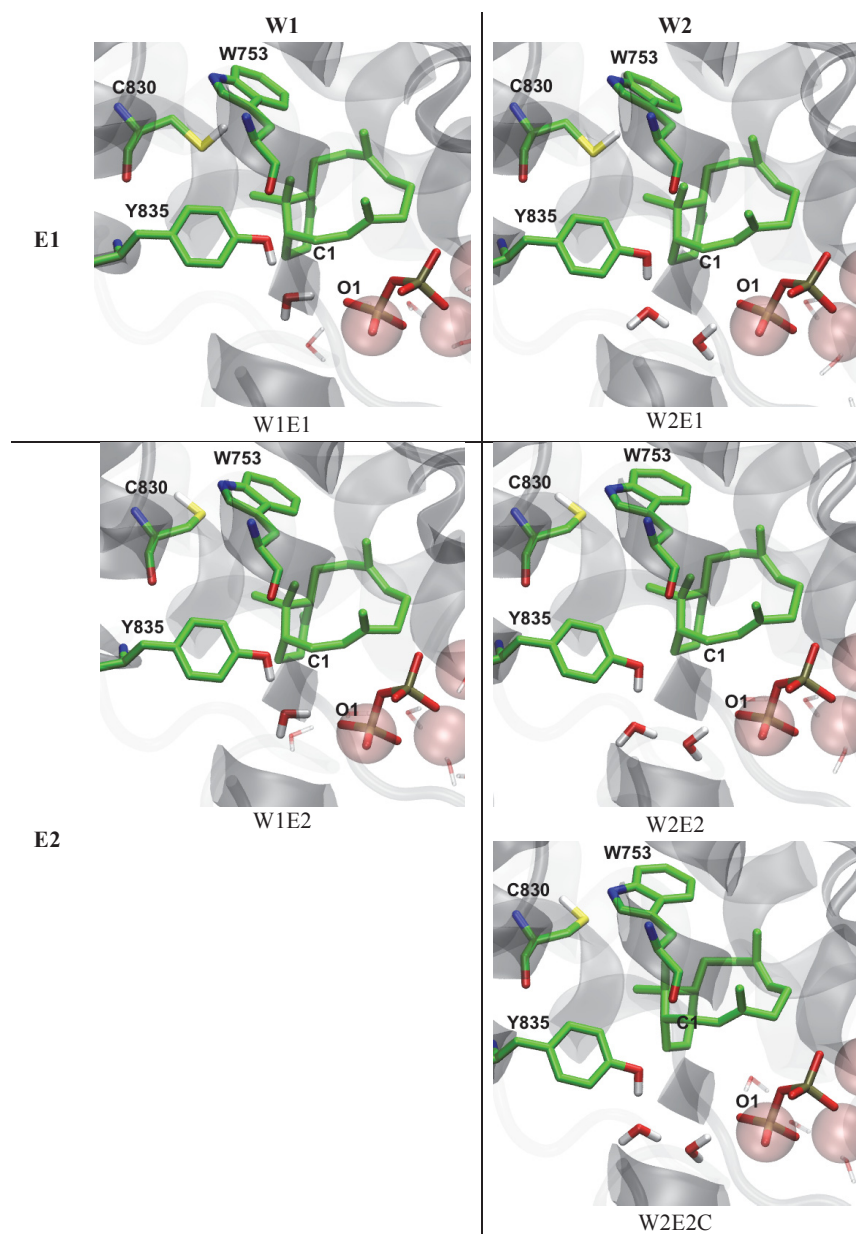
Figure S2 The five different setups **C:**WxEx taken from the MD simulation of the [TXS]C complex [1]. There are one (W1) or two (W2) water molecules between Y835 and PPi:O1. C830 interacts with W753 (E1) or C830 points away from W753 (E2). Most hydrogen atoms are omitted for clarity.

2. Results of QM study

Table S2

QM energies relative to cation **C** (ΔE in kcal/mol) for single-point calculations and for (re)optimizations of the published HT-QM structures[2] with different methods.[1]

Stationary Point	Method							
	Re-optimized structures					Single-point calculations		
	B3LYP/6-31+G(d,p)	B3LYP-D/6-31+G(d,p)	ωB97xD/6-31+G(d,p)	M06-2X/6-31+G(d,p)	M06-2X/TZVP	M06-2X/6-31+G(d,p)	mPW1PW91/6-31+G(d,p)[2]	DLPNO-CCSD(T)/def2-TZVPP[2]
A	15.5	25.6	35.3	32.5	31.4	35.9	29.3	37.3
TS(A-B)	17.7	DNC	DNC	DNC	DNC	DNC	29.3	DNC
B	12.5	17.1	22.7	21.4	21.0	22.9	19.1	25.3
TS(B-C)	14.8	19.4	DNC	21.6	21.2	DNC	23.2	24.5
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TS(C-D1)	13.2	10.2	13.3	11.4	11.5	11.3	11.0	14.0
TS(C-F)	8.6	6.5	9.3	7.4	7.8	8.0	6.5	10.2
F	0.6	0.3	1.0	0.4	0.6	1.6	0.4	2.2
TS(F-D1)	8.0	6.6	8.8	7.2	7.7	7.2	6.0	9.8
D1	2.4	1.9	2.3	1.9	2.1	2.8	2.1	3.0
TS(D1-D2)	6.0	6.5	5.9	4.9	4.9	5.1	6.5	6.6
D2	5.3	5.3	7.5	5.7	6.0	5.8	5.4	6.5
TS(D2-E)	11.9	DNC	10.3	5.9	8.6	8.2	8.2	DNC
E	11.2	10.0	6.1	5.2	6.0	5.3	5.7	1.9

[1] DNC = calculation does not converge

[2] Reported relative energies do not include ZPE corrections.

The published HT-QM structures[2] were re-optimized with different methods (see Method section in the main text and Table S1). For cation **A** the structures obtained with M06-2X/6-31+G(d,p) and M06-2X/TZVP deviated from the HT-QM structure with an RMSD of 0.37 Å for the aligned carbon atoms. For cation **B** the structure is already more similar to the HT-QM structure (RMSD of 0.12 Å) and the further the reaction progresses, the smaller the differences between the HT-QM (starting) structure and the re-optimized structure, indicated by RMSD values of 0.06 Å and less for cation **D2** and cation **E**.

Because of conformational changes during re-optimization, a connected pathway between the optimized structures of cation **A** and cation **E** is no longer guaranteed, and optimizations for the transition states TS(A-B), TS(B-C), and TS(D1-D2) do not converge at all levels of theory applied. For comparison to QM/MM data we therefore use the data from the M06-2X/6-31+G(d,p) single-point calculations. Comparison of the results for M06-2X/6-31+G(d,p) and M06-2X/TZVP shows that basis set effects are generally small.

5

Table S3

QM energies relative to cation C_{chair} (ΔE in kcal/mol) for structures of different cations optimized with the A ring (**C**, **F**, **D1**, and **D2**) and the C ring (**E** and **E2**) in chair-like and boat-like conformation.

| Cation | Conformation of A or C ring | | | |
| | Chair | | Boat | |
	M06-2X/ 6-31+G(d,p) [1]	DLPNO-CCSD(T)/ def2-TZVPP [2]	M06-2X/ 6-31+G(d,p) [1]	DLPNO-CCSD(T)/ def2-TZVPP [2]
C	0.0	0.0	-0.1[3]	NC[5]
F	0.4	2.2	0.4[3]	NC[5]
D1	1.9	3.1	7.0	6.4
D2	5.7	6.5	5.9[3]	NC[5]
E	5.2	1.9	1.5	-2.8
E2	NC[6]	NC[6]	1.4[4]	-3.0[4]

[1] Optimizations start from the geometries of the cations as found in the QM/MM calculations.
[2] Single-point calculation on M06-2X/6-31+G(d,p) optimized structures. Relative energies do not include ZPE corrections.
[3] During optimization, the structure converts back to the chair-like conformation.
[4] Species E_{C_boat} and $E2_{C\_boat}$ differ in the orientation of the cation in the binding pocket. As expected, in the gas phase, their energies are comparable.
[5] NC = not calculated; these structures convert to chair-like conformation during optimization at the M06-2X level, and hence the expensive CCSD(T) calculations were not performed.
[6] $E2_{C\_chair}$ is not observed in the QM/MM calculations.

Table S4

QM energies (ΔE in kcal/mol) for the different (side) products optimized with the A ring (products V-V2) and C ring (products T, T1) in chair-like and boat-like conformation.[1]

| Product | Conformation of A or C ring | | | |
| | Chair | | Boat | |
	M06-2X/ 6-31+G(d,p) [2]	DLPNO-CCSD(T)/ def2-TZVPP [3]	M06-2X/ 6-31+G(d,p) [2]	DLPNO-CCSD(T)/ def2-TZVPP [1]
T	54.7	54.1	0.0	0.0
T1	6.0	5.1	3.3	3.4
V	-0.4	2.2	NC[4]	NC[4]
V1	3.6	5.7	8.4	10.5
V2	0.5	2.6	4.8	7.2

[1] Energies are given relative to the main product T_{C_boat} (taxadiene with the C ring in boat-like conformation).
[2] Optimizations start from the geometries of the products as found in the QM/MM calculations.
[3] Singe-point calculations on M06-2X/6-31+G(d,p) optimized structures. Relative energies do not include ZPE corrections
[4] NC = not calculated. We did not find cation $^{TXS}C_{boat}$ which suggests that product V_{boat} may be unlikely.

6

3. GGPP → E QM/MM reaction profile
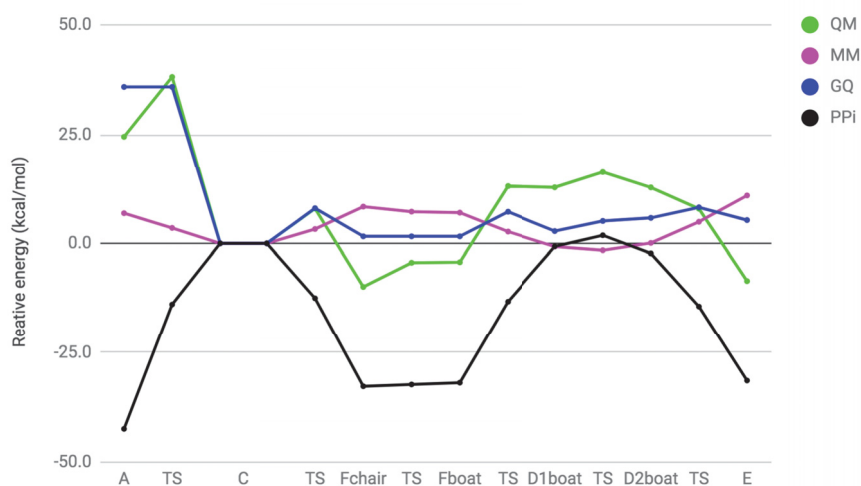
Figure S3

Partitioning of the QM/MM energy into QM (green) and MM (purple) components for setup W1E2 (see Figure S2). The QM energy component of the QM/MM energy includes the electrostatic interaction between the QM region and the MM region (electrostatic embedding, see Methods section in the main text). In black: contribution of PPi to the QM energy as calculated using an electrostatic perturbation approach[3]–[5]. In blue: the reference energies from M06-2X/6-31+G(d,p) single-point calculations on the HT-QM structures. Energies are given relative to cation **C**. A similar profile of the PPi energy contribution to the QM/MM reaction profile is expected for all other setups, since the variation of the interatomic distance between the center of the positive charge of the cations and PPi is similar along all computed reaction profiles.

7
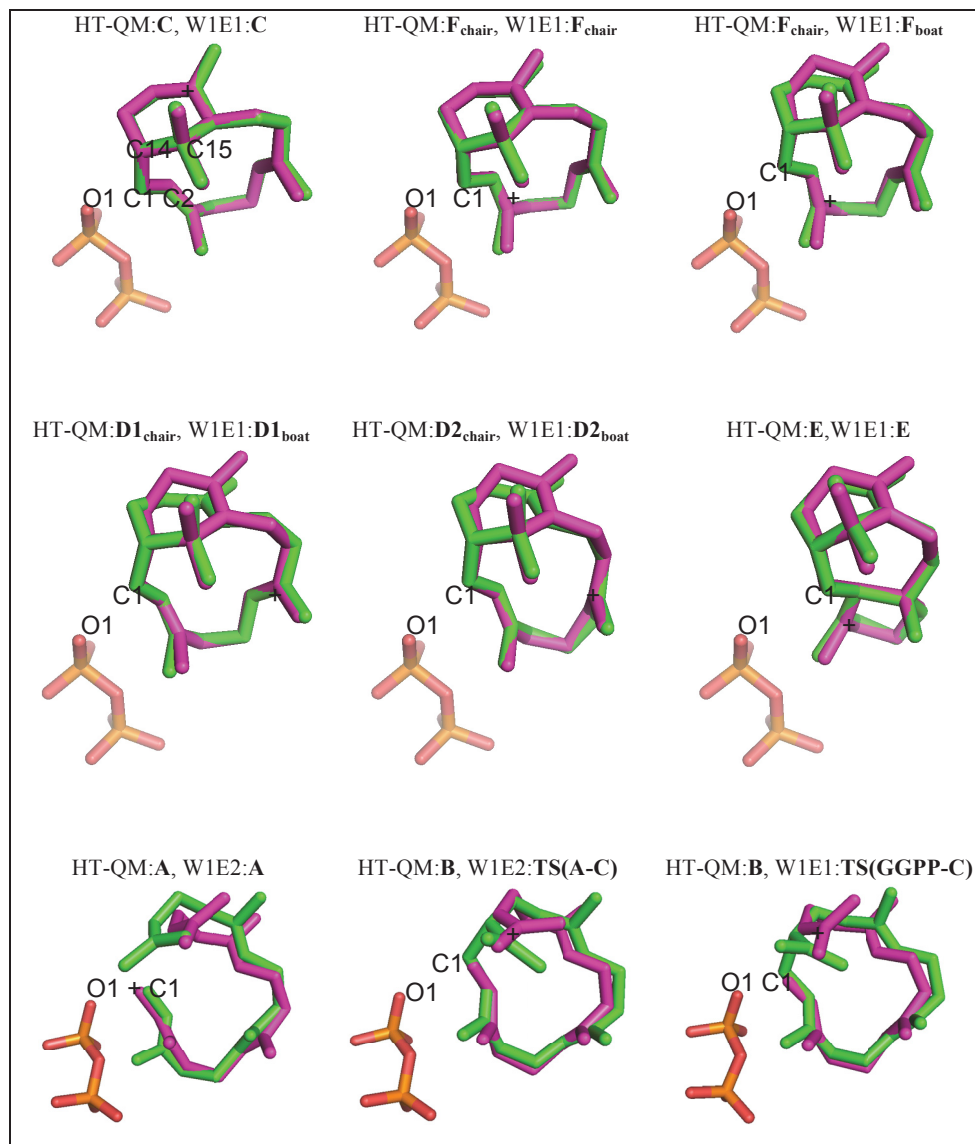
Figure S4
Overlay of HT-QM cation structures (green) and WxEx (see Figure S2) carbocation structures
(magenta). Hydrogen atoms are omitted for clarity.

8
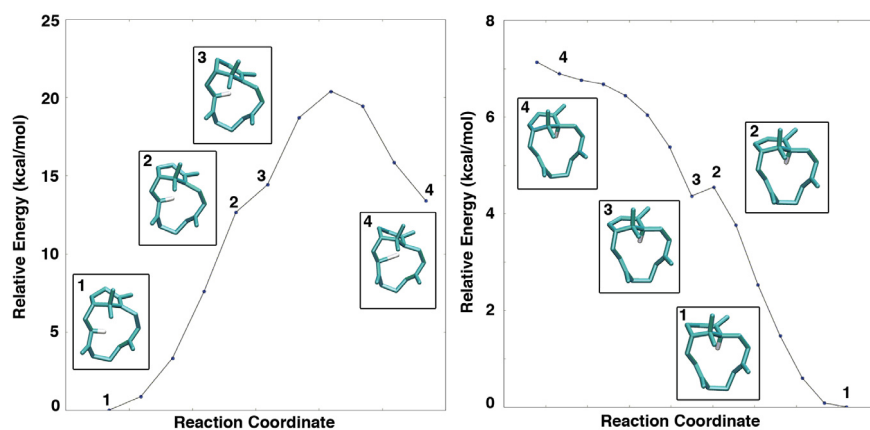
Figure S5

Left: scan from $\mathbf{F_{boat}}$ to $\mathbf{C_{boat}}$ for setup W1E1. Between points 2 and 3 the A ring changes from boat-like to chair-like conformation. Right: scan from $\mathbf{C_{chair}}$ to $\mathbf{C_{boat}}$ for setup W1E1. Though point 4 clearly is $\mathbf{C_{boat}}$, it is not a minimum and optimization leads back to $\mathbf{C_{chair}}$.
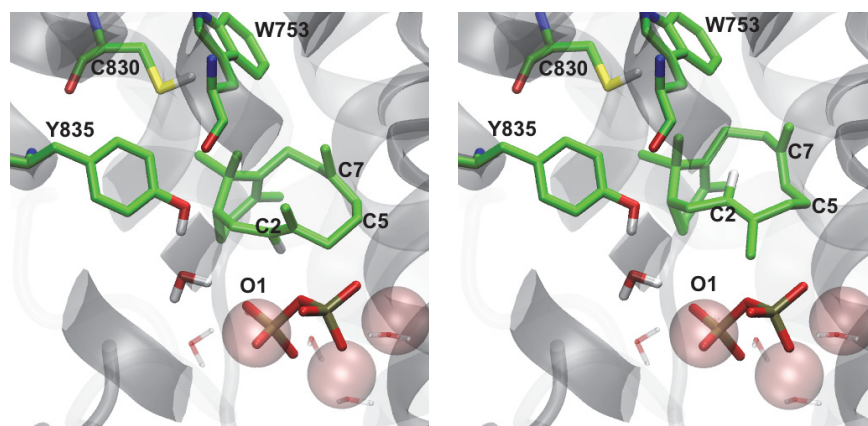
Figure S6

Left: W1E1:$\mathbf{D2_{boat}}$ as observed on the catalytic pathway (see Figure 4 of the main text). Right: W1E1:$\mathbf{D2_{boat}}$ after internal rotation around the C3-C4 bond. As can be seen, this will lead to the wrong orientation of H2 in the final product (see Figure 1 of the main text). Most hydrogen atoms are omitted for clarity.

9

Table S5

Important bond distances for positioning of the isopropyl tail of GGPP in the TXS binding pocket[1]

Distance	WxE1[2]	WxE2[3]	W1Ex[4]	W2Ex[5]
GGPP:C1 - GGPP:C14	3.6 (0.3)	3.9 (0.2)		
GGPP:C1 - GGPP:C15	4.6 (0.1)	5.0 (0.3)		
GGPP:C10 - GGPP:C15	3.1 (0.0)	3.3 (0.1)		
C830:S_C - GGPP:C15	4.1 (0.0)	5.1 (0.3)		
W753:C_W - GGPP:C15	5.3 (0.0)	5.0 (0.2)		
Y835:H_Y - GGPP:C15			3.8 (0.0)	3.3 (0.2)
F834:C_F- GGPP:C15			3.9 (0.0)	4.4 (0.2)

[1] Distances are giving in Å and correspond to average values as indicated below (standard deviations in parentheses). Labels can be found in Figure S7.
[2] Average over W1E1 and W2E1; [3] Average over W1E2, W2E2, and W2E2C; [4] Average over W1E1 and W1E2; [5] Average over W2E1, W2E2, and W2E2C.
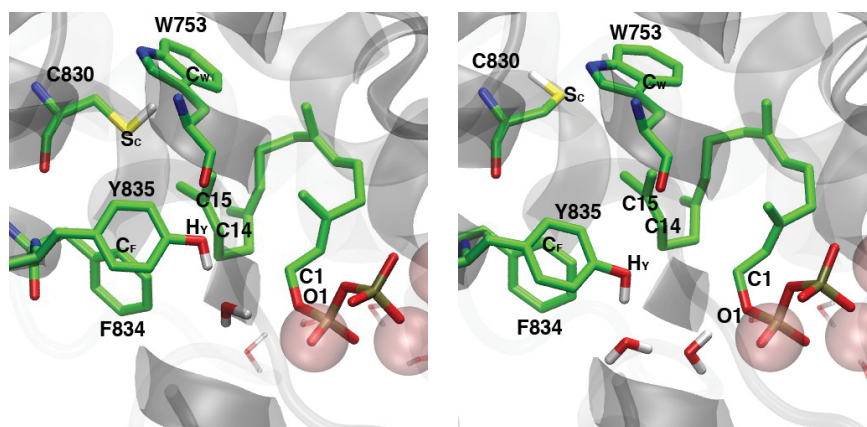
Figure S7

Left: W1E1:GGPP. Right: W2E2:GGPP. Most hydrogen atoms are omitted for clarity.
When the thiol group of C830 interacts with W753 (left), the C1-C14 and C1-C15 distances are shorter than when the thiol is flipped (right). See Table S5 above.

10
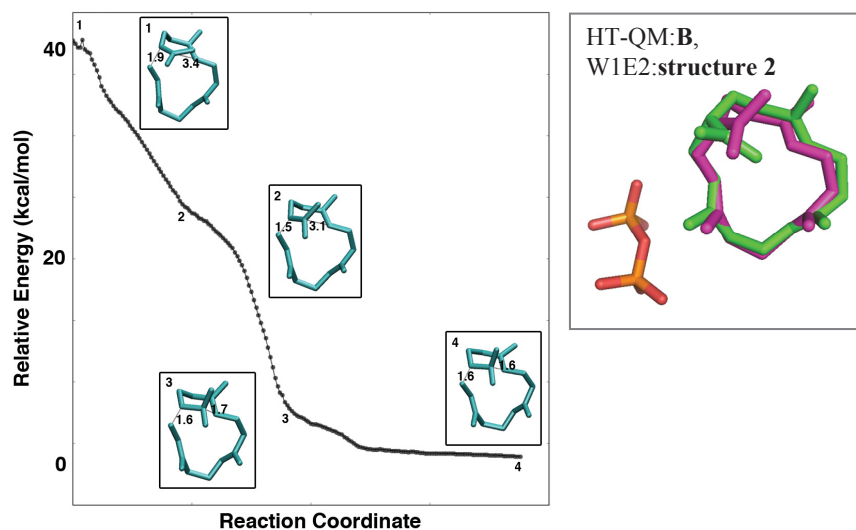
Figure S8

Optimization of W1E2:**TS(A-C)** passes through a structure similar to HT-QM cation **B**. Left: energy profile of IRC-like calculation from W1E2:**TS(A-C)** to W1E2:**C**. Right: Overlay of HT-QM cation **B** (green) and structure 2 from left scan (magenta). Hydrogen atoms are omitted for clarity.

11

Table S6

Average QM/MM energies (in kcal/mol with respect to the ^{TXS}C complex) of the computed energy profiles for the conversion of GGPP to **T** in the TXS environment. Also shown is the partitioning of the QM/MM energy into QM and MM components. Standard deviations are given in parentheses.

System[1]	Stationary point	QM	MM	QM/MM
WxE1[2]	GGPP	11.3 (9.8)	2.5 (6.4)	13.8 (3.5)
	TS(GGPP-C)	36.5 (3.8)	3.6 (0.1)	40.1 (3.9)
WxE2[3]	GGPP	15.1 (3.8)	1.6 (2.7)	16.8 (3.8)
	TS(GGPP-A)	33.8 (6.0)	-0.5 (6.6)	33.3 (1.7)
	A	24.0 (0.6)	6.6 (0.4)	30.7 (0.6)
	TS(A-C)	39.0 (1.3)	2.8 (1.0)	41.8 (1.9)
WxEx[4]	C	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
	TS(C-F_{chair})	7.5 (2.2)	3.7 (0.8)	11.2 (2.5)
	TS(C-$D1_{chair}$)	17.9 (4.0)	2.1 (4.1)	19.9 (1.3)
	$D1_{chair}$	5.0 (3.6)	2.7 (4.1)	7.6 (1.9)
	F_{chair}	-14.0 (3.0)	9.0 (1.7)	-5.0 (3.4)
	TS(F_{chair}-F_{boat})	-10.4 (4.2)	9.7 (2.3)	-0.7 (3.1)
	TS(F_{chair}-$D1_{chair}$)	4.3 (2.6)	5.4 (2.4)	9.6 (2.9)
	F_{boat}	-10.6 (4.0)	9.0 (2.0)	-1.6 (3.3)
	TS(F_{boat}-$D1_{boat}$)	7.3 (3.7)	4.8 (1.3)	12.1 (2.9)
	$D1_{boat}$	7.2 (3.8)	3.1 (2.3)	10.3 (2.4)
	TS($D1_{boat}$-$D2_{boat}$)	8.5 (5.0)	4.8 (4.2)	13.3 (1.9)
	$D2_{boat}$	4.9 (6.2)	5.9 (3.8)	10.8 (2.9)
	TS($D2_{boat}$-E_{C_chair})	2.0 (6.5)	8.6 (4.2)	10.6 (3.9)
	E_{C_chair}	-15.3 (7.3)	13.5 (6.2)	-1.7 (2.8)

[1] The pathway from ^{TXS}GGPP to ^{TXS}C contains one or two steps depending on the NTRC configuration and is therefore split for WxE1 and WxE2. See main text, section 3.1, Other setups.
[2] Averaged over W1E1 and W2E1.
[3] Averaged over W1E2, W2E2, and W2E2C.
[4] Averaged over W1E1,W2E1, W1E2, W2E2, and W2E2C.

Table S7

Average barriers and reaction energies (in kcal/mol) for the conformational and translational changes of cation **E** along the computed reaction profiles of the conversion of GGPP to **T** in the TXS environment. Standard deviations are given in parentheses.

Reaction step	System	Reaction energy	Reaction barrier
$E_{C\_chair} \rightarrow E_{C\_boat}$	WxE1[1]	1.2 (0.2)	19.2 (1.5)
	WxE2[2]	0.3 (0.4)	9.3 (1.7)
$E_{C\_boat} \rightarrow E2_{C\_boat}$	WxEx[3]	6.0 (7.3)	8.7 (5.6)

[1] Averaged over W1E1 and W2E1.
[2] Averaged over W1E2, W2E2 and W2E2C.
[3] Averaged over W1E1, W1E2 and W2E1. For W2E2 and W2E2C the $E_{C\_boat} \rightarrow E2_{C\_boat}$ conversion was not found

12

Table S8

Important distances (in Å) for positioning (top) and deprotonation pathways (bottom) of the different conformations of cation **E** in the TXS binding pocket.[1]

		Distance (Å)				
Conformation and setup		$C830{:}S_C$ - $W753{:}C_W$	$C830{:}S_C$ - C15	$W753{:}C_W$- C15	$W753{:}C_W$- C16	PPi:O1- C20
E_{C_chair}	WxE1[2]	3.8 (0.2)	4.5 (0.1)	5.6 (0.1)	4.2 (0.1)	3.3 (0.2)
	W2E2x[3]	5.1 (0.1)	5.1 (0.2)	4.9 (0.0)	3.4 (0.0)	4.6 (0.0)
	W1E2	5.0	4.8	5.7	4.2	4.1
E_{C_boat}	WxE1[2]	3.8 (0.2)	4.5 (0.0)	5.6 (0.1)	4.1 (0.0)	3.1 (0.1)
	W2E2x[3]	5.2 (0.0)	4.8 (0.1)	4.6 (0.0)	3.6 (0.1)	4.2 (0.3)
	W1E2	5.1	4.8	5.7	4.2	3.0
$E2_{C\_boat}$	WxE1[2]	3.7 (0.2)	4.3 (0.2)	5.2 (0.5)	4.2 (0.5)	3.6 (0.6)
	W1E2	5.1	4.8	5.3	4.2	4.1

		Distance (Å)				
Conformation and setup		PPi:O1- H20[4]	PPi:O1- H4α	PPi:O1- H4β	PPi:O4- H4α	PPi:O4- H4β
E_{C_chair}	WxE1[2]	2.3 (0.3)	4.1 (0.5)	3.8 (0.0)	5.2 (0.6)	3.4 (0.5)
	W2E2x[3]	3.7 (0.1)	4.1 (0.2)	4.3 (0.1)	4.3 (0.1)	2.5 (0.1)
	W1E2	3.0	5.2	4.5	5.3	3.5
E_{C_boat}	WxE1[2]	2.3 (0.4)	4.4 (0.3)	5.0 (0.1)	3.0 (0.2)	4.4 (0.4)
	W2E2x[3]	3.1 (0.3)	4.7 (0.5)	5.9 (0.6)	2.6 (0.4)	2.8 (0.2)
	W1E2	2.6	4.9	4.4	4.4	2.9
$E2_{C\_boat}$	WxE1[2]	2.8 (0.8)	3.1 (0.1)	2.2 (0.1)	4.3 (0.0)	4.2 (0.2)
	W1E2	3.4	2.6	2.9	4.2	4.3

[1] Distances correspond to average values as indicated below. Standard deviations are given in parentheses. See Figs. S9-S10 for atom labels. [2] Average over W1E1 and W2E1. [3] Average over W2E2 and W2E2C. [4] Hydrogen atom closest to base (PPi:O1).
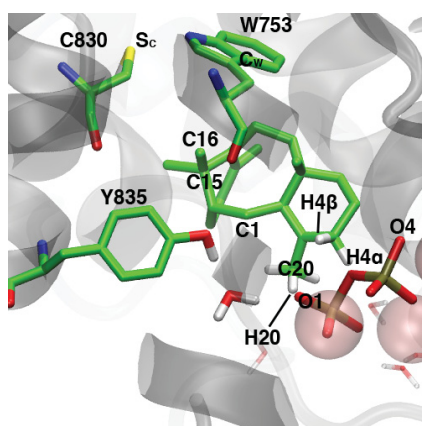
Figure S9
The $^{TXS}E2_{C\_boat}$ complex from snapshot W1E2E showing all relevant labels for Table S8.

13
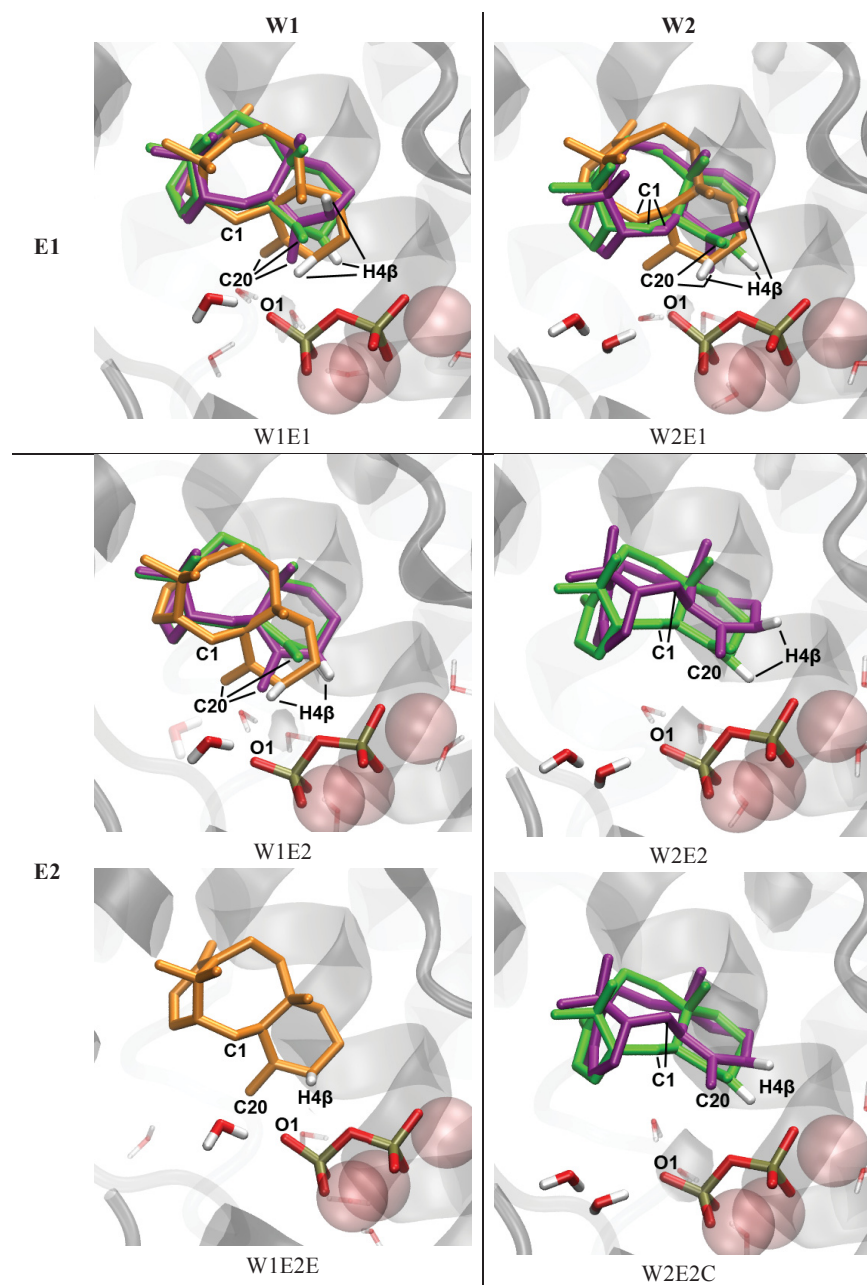
Figure S10

^{TXS}**E** complexes identified in the QM/MM calculations. ^{TXS}**E**$_{C\_chair}$ (cation in green) originates from the ^{TXS}**D2**$_{boat}$ complex. ^{TXS}**E**$_{C\_boat}$ (cation in purple) results from a conformational change of the C ring to boat-like conformation. ^{TXS}**E2**$_{C\_boat}$ (cation in orange) originates from a slight rotation of the **E**$_{C\_boat}$ conformer. All hydrogen atoms except for H4β are omitted for clarity. From the MD simulations of ^{TXS}**E**[1], an additional snapshot of ^{TXS}**E2**$_{C\_boat}$ was taken, W1E2E. The other snapshots correspond to setups described in Fig. S2.

14

Table S9

QM/MM energies (in kcal/mol) of the reaction profile computed for the conversion of **C** to **T** using the $^{TXS}E2_{C\_boat}$ snapshot, W1E2E. For comparison, the data reported by Ansbacher et al.[6] using the FHM is shown. Energies are given relative to the ^{TXS}C complex[1].

	$^{TXS}E2_{C\_boat}$ snapshot, W1E2E			Ansbacher et. al.[6]
	QM	**MM**	**QM/MM**	
C	0.0	0.0	0.0	0.0
TS(C-F$_{boat}$)	8.6	2.0	10.6	4.9
TS(C-D1$_{boat}$)	6.4	2.8	9.2	4.0
F$_{boat}$	-15.2	5.7	-9.6	-9.9
TS(F$_{boat}$ -D1$_{boat}$)	4.9	3.6	8.5	6.6
D1$_{boat}$	-8.4	3.9	-4.6	-15.8
D2$_{boat}$	3.8	-2.9	0.9	Not reported[2]
TS(D2$_{boat}$ -E2$_{C\ boat}$)	6.0	0.0	6.0	-11.0
E2$_{C\ boat}$	-22.1	-6.7	-28.8	-24.4
TS(E2$_{C\ boat}$ -T)	-21.0	-5.1	-26.1	-22.1
T	-51.5	6.0	-45.6	-34.8
TS(E2$_{C\ boat}$ -T1)	-14.6	-8.0	-22.6	Not reported[2]
T1	-38.9	0.5	-38.4	Not reported[2]

[1] The ^{TXS}C complex found by back propagation for W1E2E has a different orientation in the binding pocket than the setups taken from the MD simulation of the ^{TXS}C complex[1] (see Table S10).

[2] An energy value for this complex is not reported in [6].

Table S10.

Distance (Å) between the center of positive charge of cations **C**, **D1** and **F** (CX^+)[1] and surrounding atoms (important for charge stabilization) in the QM/MM reaction profiles obtained for setups W1E1 and W1E2E.

Interatomic distance[1]	System					
	W1E1			W1E2E		
	C	D1$_{boat}$	F$_{boat}$	C	D1$_{boat}$	F$_{boat}$
CX^+-PPi:O1	6.2	7.4	3.5	7.8	5.1	4.1
CX^+-PPi:O2	7.5	7.4	3.6	8.6	5.2	5.6
CX^+-PPi:O4	9.2	8.0	5.0	9.8	6.1	7.6
CX^+-D613$_{OD1}$	9.3	7.6	4.8	9.3	5.9	7.1
CX^+-D613$_{OD2}$	8.3	7.7	5.3	8.8	6.0	5.9
CX^+-W753[2]	7.4	6.2	8.0	6.2	5.8	8.9
CX^+-Y841[2]	5.8	9.7	9.4	7.7	9.7	7.1

[1] X=11 for cation **C**; X=7 for cation **D1** and X=3 for cation **F**. Labels can be found in Fig. 3 of the main text and Figs. S12b and S12d.

[2] For W753 and Y841 we report the distance between CX^+ and the centroid of the heavy ring atoms.

15

4. Deprotonation data

Table S11

Average barriers and reaction energies (in kcal/mol) for deprotonation of C4 on the α face or β face by PPi:O1 or PPi:O4 for the different conformations of the ^{TXS}E complex. Standard deviations are given in parentheses.[1]

Reaction	Reacting atoms[2]	Reaction energy	Reaction barrier
$E_{C\_chair} \rightarrow T$	O1-H4β[3]	4.1 (6.8)	53.6 (11.6)
	O4-H4α[4]	0.9 (8.4)	18.0 (7.7)
	O4-H4β[5]	NF[8]	NF[8]
$E_{C\_boat} \rightarrow T$	O4-H4α[6]	-10.5 (5.0)	7.4 (3.4)
	O4-H4β[5]	-3.9 (8.1)	20.8 (6.9)
$E2_{C\_boat} \rightarrow T$	O1-H4β[3]	-13.4 (3.7)	2.7 (1.7)
$E_{C\_chair} \rightarrow T1$	O1- C20:H[7,5]	-7.4 (3.2)	8.6 (5.9)
$E_{C\_boat} \rightarrow T1$	O1- C20:H[7,5]	-7.6 (4.1)	8.8 (6.0)

[1] Results are only reported for the setups for which the indicated reaction was found. These setups are specified in the footnotes below. [2] Atom labels can be found in Figures 1 and 3 of the main text.
[3] Average over W1E1, W1E2, and W2E1. [4] Average over W1E2, W2E1 and W2E2C.
[5] Average over W1E1,W2E1, W1E2, W2E2, and W2E2C. [6] Average over W2E2 and W2E2C.
[7] Hydrogen atom closest to base (PPi:O1).
[8] NF = not found; scan does not pass a transition state, but continues to rise in energy (like Fig. S5 right).

Table S12

Average barriers and reaction energies (in kcal/mol) of the expected preferred deprotonation pathway of ^{TXS}C (a water-assisted proton transfer to PPi yielding **V**) and of ^{TXS}F (deprotonation either directly by PPi or by a water-assisted proton transfer to PPi to produce either **V1** or **V2**). Standard deviations are given in parentheses.[1]

Reaction	Reacting atoms[2]	Reaction energy	Reaction barrier
$C \rightarrow V$	C12:H[3]-O$_w$-O1[4]	-23.4 (2.4)	5.6 (3.0)
$F_{chair} \rightarrow V1$	C20:H[3]-O1[5]	-10.7 (2.8)	4.7 (1.3)
$F_{boat} \rightarrow V1$	C20:H[3]-O1[6]	-14.4 (0.8)	3.9 (1.5)
$F_{boat} \rightarrow V2$	C2:H[3]-O1[7]	-18.2	0.0
$F_{boat} \rightarrow V2$	C2:H[3]-O$_w$-O1[8]	-18.0	1.2
$F_{chair} \rightarrow V2$	C2:H[3]-O$_w$-O1[9]	-18.4 (2.5)	2.5 (1.5)

[1] Results are only reported for the setups for which the indicated reaction was found. These setups are specified in the footnotes below. [2] Atom labels can be found in Figures 1 and 3 of the main text.
[3] Hydrogen atom closest to base (PPi:O1). [4] Average over W1E1,W2E1, W1E2, W2E2, and W2E2C.
[5] Average over W2E1, W2E1F1 and W2E1F2 (two new snapshots of cation **F** taken from MD of ^{TXS}**F**). New snapshots were selected based on cation **F** configuration and the orientation of C20:H and C2:H with respect to O1 to sample all the deprotonation pathways observed during the MD simulation[1]. They are numbered from F1 to F6 and labeled using standard naming conventions for the water network and NTRC orientation, e.g. W2E1F1 has two water molecules between Y835 and PPi:O1 and NTRC orientation E1.
[6] Average over W1E1 and W1E1F3. [7] Data for setup W2E1F4. [8] Data for setup W1E2F5.
[9] Average over W1E1, W1E2, W2E1, W2E2 and W2E2F6.

Table S13

Based on the computed distances, the direct and water-mediated deprotonation by PPi:O1 to side products **V**, **V1** and **V2** seem unlikely for structures on the pathway from the ^{TXS}E2$_{C\_boat}$ setup W1E2E.

Proton acceptor-Deprotonation target	Distance (Å)	Water bridge
PPi:O1-C:C12	7.1	No water molecules within a 4 Å radius
PPi:O1-F$_{boat}$:C2	5.5	No water molecules within a 4 Å radius
PPi:O1-F$_{boat}$:C20	4.3	Water bridge between C20:H and PPi but to less basic oxygen (O13)

16

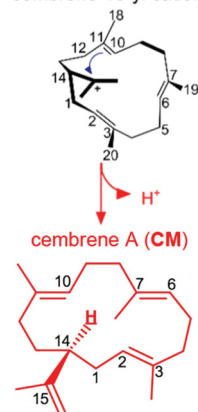**Figure S11** Cembrene A (CM) is a side product observed in the product distribution for mutated TXS. However, in the wild-type product distribution, CM is not detected.[7]

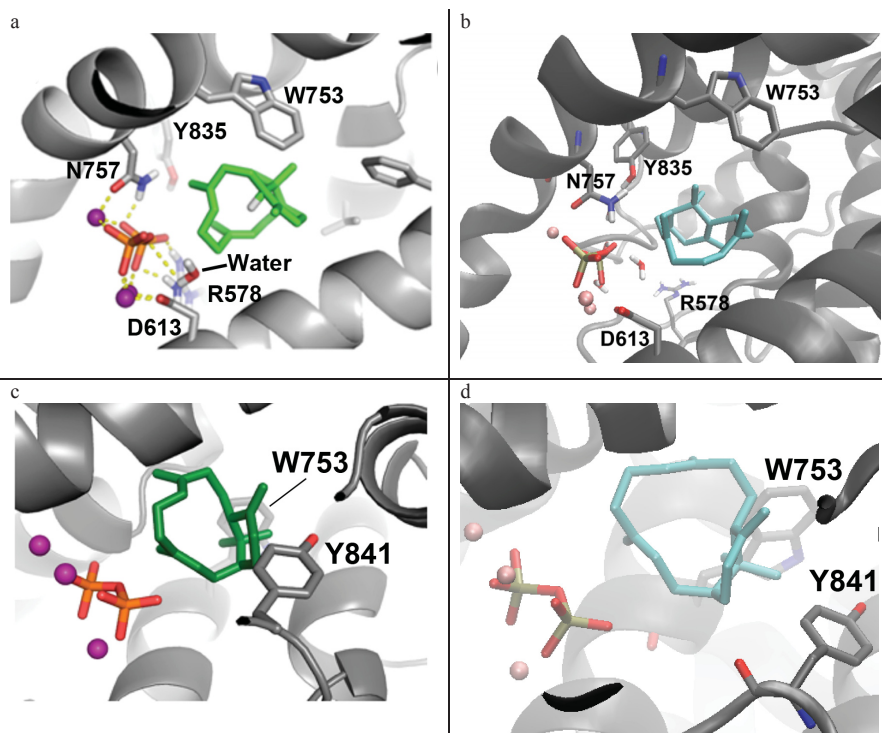**5. Differences between the SHM and the FHM**

Figure S12 Plots a) and c) correspond to figure 3a from Freud et. al.[8] and 2c from Ansbacher et al.[6] respectively; residue labels have been added. Plots b) and d) depict setup W1E1:**C**; structures have been rotated to facilitate comparison with a) and c). Plots b) and d) highlight the same residues as plots a) and c). In d) water molecules are omitted for clarity.

Number of water molecules
While the setup with the FHM (Fig. S12a) contains a single active-site water molecule, the setup with the SHM (Fig. S12b) has 4 water molecules within 3 Å of the carbocation.

Orientation of the substrate in the binding pocket
The orientation of cation **C** in the binding pocket is very different in Figs. S12a and S12b, while the location and orientation of cation **C** in the binding pocket is similar in Figs. S12c and S12d. Since Figs. S12b and S12d show the same setup, the orientation of cation **C** in Figs. S12a and S12c must be different in the two published studies.[6,8] The differences between our setup and the setup in [6] are significantly smaller than those with setup [8], but they are not negligible. For example, in the FHM from [6] the positive center at ^{TXs}**C**:C11 forms a π−cation interaction with Y841[6] (no distance reported), while in the SHM the C11$^+$···Y841 distance is 5.8 Å (see Table S10), so the π−cation interaction is very weak.

Structure of binding pocket
Both R578 (A-C loop) and D839 (J-K loop) are located over 12.5 Å away from PPi in the SHM, while the former residue interacts directly with the PPi moiety in the FHM and the latter makes a water-mediated hydrogen bond with PPi. Therefore it appears that the A–C (G570–H579) and J–K (F837–E846) loops are positioned differently leading to different active-site architectures.

18

6. References

[1] A. M. Escorcia, J. P. M. van Rijn, G. J. Cheng, P. Schrepfer, T. B. Brück, W. Thiel, *J. Comput. Chem.* **2018**, *39*, 1215–1225

[2] Y. J. Hong, D. J. Tantillo, *J. Am. Chem. Soc.* **2011**, *133*, 18249–18256.

[3] A. M. Escorcia, K. Sen, M. C. Daza, M. Doerr, W. Thiel, *ACS Catal.* **2017**, *7*, 115–127.

[4] P.A. Bash, M. J. Field, R. C. Davenport, G. A. Petsko, D. Ringe, M. Karplus, *Biochemistry*, **1991**, *30*, 5826−583.

[5] J. Gao, X. Xia, *Science*, **1992**, *258*, 631–635.

[6] T. Ansbacher, Y. Freud, D. T. Major, *Biochemistry*, **2018**, *57*, 3773–3779.

[7] P. Schrepfer, A. Buettner, C. Goerner, M. Hertel, J. van Rijn, F. Wallrapp, W. Eisenreich, V. Sieber, R. Kourist, T. Brück, *Proc. Natl. Acad. Sci.* **2016**, *113*, E958-967.

[8] Y. Freud, T. Ansbacher, D. T. Major, *ACS Catal.* **2017**, *7*, 7653–7657.

Reference for Gaussian09:

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01; Gaussian Inc., Wallingford, CT, 2013.

19

C. Effects of the Acyl Moiety on CalB-catalyzed Hydrolysis of Propranolol Esters.

Gui-Juan Cheng, Andrés M. Escorcia, **Jeaphianne P. M. van Rijn** and Walter Thiel

I carried out part of the QM/MM calculations. I wrote the scripts to determine the productive Michaelis complexes and for analysis of the bond distances. I helped interpret the results and contributed to the writing of the manuscript.

Appendix C
Effects of the Acyl Moiety on CalB-catalyzed Hydrolysis of Propanolol Esters

Gui-Juan Cheng, Andrés M. Escorcia, Jeaphianne P. M. van Rijn and Walter Thiel*

*W.T.: Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim, Germany. Email: thiel@kofo.mpg.de

Content:

1. Docking and post-docking optimization
    (Figures 2-4)

2. Molecular dynamic studies of Michealis complexes of CalB-propranolol esters
    (Figures 5-7, Tables 1-3)

3. QM/MM study of hydrolysis reaction of O-acetyl-propranolol (**M0**)
    (Figures 8-10, Tables 4-9)
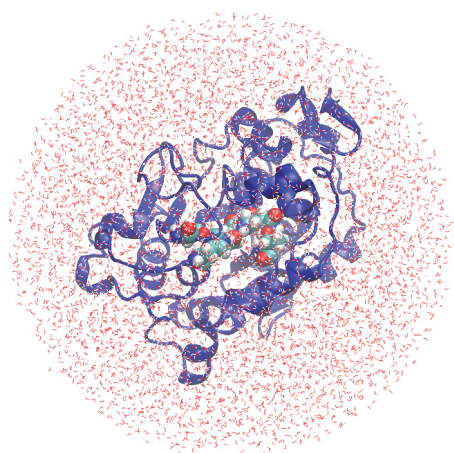
Figure 1. Representative structure of solvated CalB (blue). The residues of the catalytic triad and the oxyanion hole are shown in sphere representation and colored according to atom types.

1. Docking and post-docking optimization
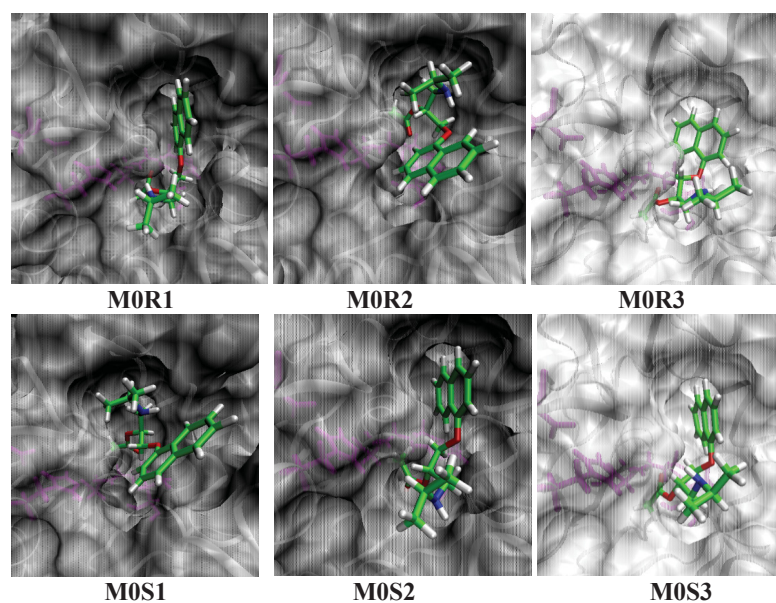
Figure 2. The productive MCCs of **M0** after post-docking optimization. Both **M0R1** and **M0R3** adopt binding mode Ia, but the naphthoxy group pointing toward the exterior and interior of the binding pocket in **M0R1** and **M0R3**, respectively. Complex **M0R2** adopts binding mode IIb. **M0S2** and **M0S3** adopt binding modes Ib and Ia,. **M0S1** is binding mode IIb.
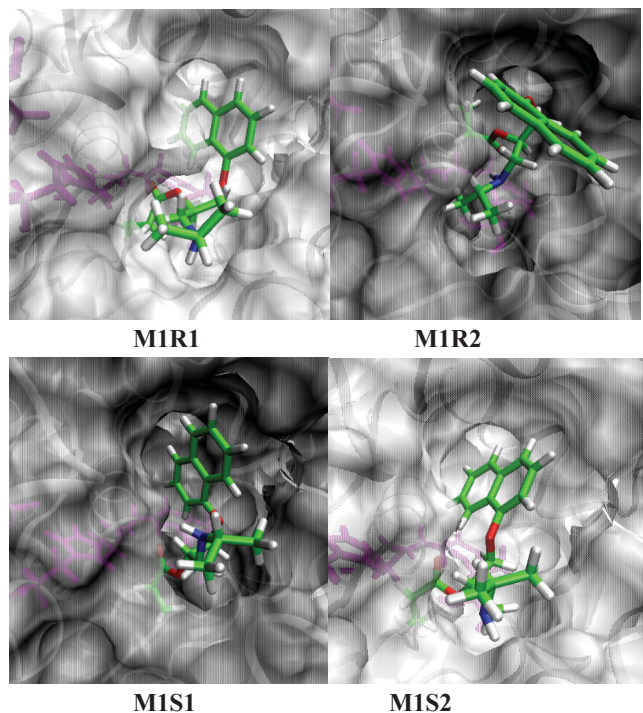
Figure 3. Three productive MCCs were identified for **M1** after minimization and we also included **M1S1** for MD study because d3 is 4.4 Å and d5 is 3.2 Å indicating a weak hydrogen bonding interaction between M1 and oxyanion hole and **M1S1** may form better interactions with the oxyanion hole during the MD simulation. **M1R1** possesses binding mode Ia. Its acyl group points downward and has an extended linear conformation. **M1R2** adopts binding mode Ib: its acyl group point upward and the amine group of **M1** forms hydrogen bond with Thr40. Both **M1S1** and **M1S2** adopt binding mode Ia and their acyl group forms a compact conformation. The difference between them lies in the orientation of naphthoxy group.
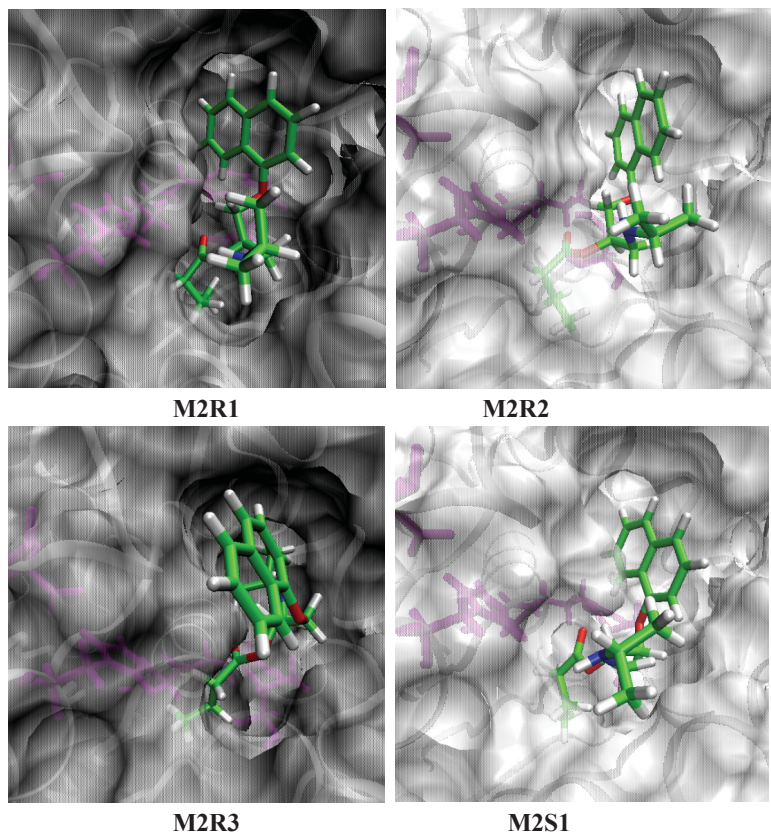
Figure 4. Three CalB-M2 complexes were characterized as productive MCCs after post-docking minimization and we also included **M2S1** for MD study because the d3 is 4.6 Å and d5 is 3.6 Å and **M2S1** may accommodate better to the binding pocket during the MD simulation. **M2R1** and **M2R2** adopt binding mode Ia and their acyl groups adopt a compact conformation. The naphthoxy group points toward to the interior and exterior of the binding pocket in **M2R1** and **M2R2**, respectively. In **M2R3**, the isopropylamine group occupies the large binding pocket and the naphthoxy group occupies the entrance of the binding pocket. The propyl chain of acyl group for **M2R3** adopts a linear and extended conformation. The substrate binds with CalB in binding mode Ia in **M2S1**.

2. Molecular dynamic studies of Michealis complexes of CalB-propranolol esters
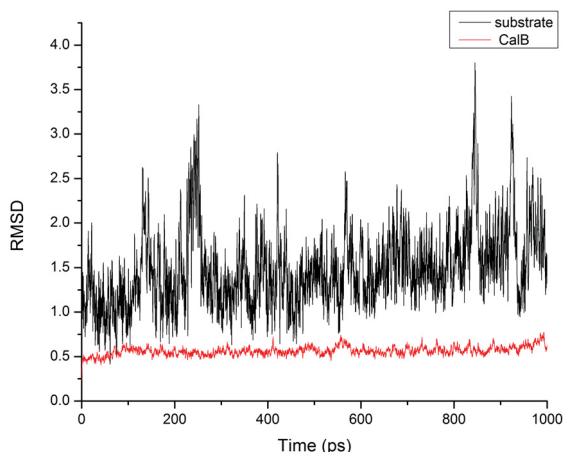
Figure 5. The root mean square deviations (RMSD) of the heavy atoms of protein backbone (red) and RMSD of substrate (black) during the MD1 simulation of **M0R2**. The MD simulations of other MCCs except **M0R1** have similar RMSD values.

Table 1. Average and deviation for distance **d3**, **d5** and **d6**[a] as well as the life time for productive conformations of **M0**.

Binding mode I[c]	Distance (Å)						Life Time (ps)	
	d3	d5	d6	d3*[b]	d5*	d6*	MD1	MD2
M0R1b	3.76(0.27)	2.58(0.48)	1.90(0.19)	3.72(0.26)	2.45(0.50)	2.00(0.26)	215.1	202.0
M0R1c	3.86(0.27)	2.63(0.48)	2.04(0.28)	3.77(0.24)	2.71(0.40)	1.89(0.18)	315.8	31.4
M0R1d	3.75(0.28)	2.53(0.47)	1.91(0.23)	3.85(0.26)	2.65(0.49)	2.23(0.31)	226.4	155.9
M0R1b'	-	-	-	3.81(0.26)	2.70(0.50)	2.03(0.30)	0.0	83.6
M0R1c'	-	-	-	4.02(0.25)	2.94(0.43)	2.26(0.32)	0.0	172.2
M0R1d'	-	-	-	3.85(0.26)	2.65(0.49)	2.23(0.31)	0.0	109.1
M0R3c	-	-	-	3.93(0.25)	2.78(0.46)	2.64(0.47)	0.0	127.5
M0R3d	-	-	-	4.06(0.31)	3.38(1.07)	3.43(1.16)	0.0	139.0
M0S2d'	4.27(0.28)	3.66(0.46)	3.84(0.40)	-	-	-	402.3	0.0
M0S2d	-	-	-	4.28(0.19)	4.34(0.60)	2.32(0.53)	0.0	103.5
M0S3d	4.14(0.23)	3.64(0.47)	3.71(0.42)	4.34(0.15)	5.90(0.84)	5.41(0.62)	267.9	54.5
Binding mode II[c]								
M0R2b	3.86(0.24)	2.56(0.71)	4.43(0.36)	-	-	-	184.6	0.0
M0R2c	4.01(0.24)	2.77(0.68)	4.66(0.39)	4.06(0.25)	2.15(0.48)	4.39(0.30)	114.3	47.1
M0R2d	3.97(0.50)	2.60(0.69)	4.57(0.35)	4.02(0.25)	2.12(0.46)	4.41(0.32)	425.1	276
M0S1a	4.39(0.05)	1.78(0.10)	4.15(0.38)	4.15(0.26)	1.84(0.33)	4.41(0.34)	0.3	12.2
M0S1b	4.40(0.10)	1.74(0.14)	4.20(0.36)	4.18(0.21)	1.84(0.32)	4.45(0.31)	3.3	69.8
M0S1c	4.22(0.20)	1.77(0.13)	4.53(0.25)	4.14(0.26)	1.78(0.14)	4.42(0.35)	50.4	454
M0S1d	4.30(0.15)	1.84(0.17)	4.62(0.23)	4.23(0.22)	1.78(0.13)	4.51(0.31)	2.5	21.5
M0S1c'	4.12(0.26)	1.76(0.13)	4.54(0.34)	-	-	-	209.1	0.0

[a] Distances are given in Å. Numbers in parentheses correspond to standard deviations from average values. See method section of the main text and Figure 3 for the definitions of distances d3, d5, and d6.
[b] Two MD simulations were performed with different initial velocity distribution as indicated by *.
[c] Binding modes indicated with ' correspond to MCC complexes in which the protein adopts an "open" conformation.

Table 2 Average and deviation for distance **d3**, **d5** and **d6**[a] as well as the life time for productive conformations of **M1**.

Binding mode I[c]	Distance (Å)						Life Time (ps)	
	d3	d5	d6	d3*[b]	d5*	d6*	MD1	MD2
M1R1d	4.41(0.10)	3.65(0.52)	3.96(0.34)	4.41(0.09)	3.76(0.43)	4.07(0.43)	13.6	18.7
M1R1d'	4.11(0.20)	3.55(0.72)	4.06(0.40)	4.21(0.19)	3.56(0.49)	4.22(0.43)	1.8	10.9
M1R2d	4.14(0.32)	1.75(0.12)	4.34(0.26)	4.35(0.20)	1.76(0.12)	4.09(0.27)	41.6	1.9
M1R2d'	4.08(0.34)	1.72(0.11)	4.25(0.28)	-	-	-	215.1	0.0
M1S1b	4.36(0.13)	3.92(0.32)	3.72(0.39)	4.37(0.14)	3.93(0.30)	3.80(0.36)	23.8	8.6
M1S1b'	4.13(0.24)	3.93(0.46)	3.83(0.55)	4.13(0.23)	3.78(0.32)	3.51(0.39)	18.9	26.9
M1S1d	4.36(0.13)	3.92(0.32)	3.68(0.37)	4.38(0.15)	3.95(0.39)	3.95(0.39)	10.6	2.8
M1S2d	4.41(0.10)	3.65(0.52)	3.96(0.34)	4.42(0.07)	3.93(0.46)	3.93(0.38)	13.6	17.5

[a] Distances are given in Å. Numbers in parentheses correspond to standard deviations from average values. See method section of the main text and Figure 3 for the definitions of distances d3, d5, and d6.
[b] Two MD simulations were performed with different initial velocity distribution as indicated by *.
[c] Binding modes indicated with ' correspond to a compact conformation while the others are extended.

Table 3 Average and deviation for distance **d3**, **d5** and **d6**[a] as well as the life time for productive conformations of **M2**.

Binding mode I	Distance (Å)						Life Time (ps)	
	d3	d5	d6	d3*[b]	d5*	d6*	MD1	MD2
M2R2a	3.98(0.28)	2.44(0.51)	3.00(0.68)	3.99(0.24)	2.34(0.41)	2.75(0.54)	80.4	153.4
M2R2b	3.87(0.37)	2.47(0.58)	3.00(0.83)	3.98(0.23)	2.34(0.40)	2.75(0.52)	2.9	34.7
M2R2c	3.88(0.28)	2.89(0.51)	2.74(0.43)	3.83(0.30)	2.35(0.44)	2.50(0.48)	274.9	411.8
M2R2d	3.90(0.30)	2.92(0.49)	2.78(0.53)	3.66(0.33)	2.12(0.35)	2.39(0.60)	32.2	158.5
M2S1d	-	-	-	4.35(0.13)	4.62(0.96)	4.88(0.64)	0	31.0

[a] Distances are given in Å. Numbers in parentheses correspond to standard deviations from average values. See method section of the main text and Figure 3 for the definitions of distances d3, d5, and d6.
[b] Two MD simulations were performed with different initial velocity distribution as indicated by *.
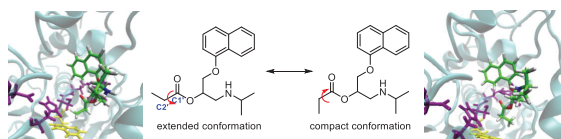
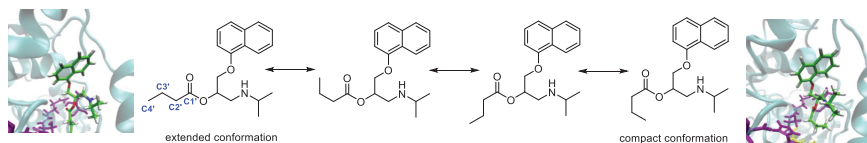**Figure 6.** Extended and compact conformation for **M1R1**.

Figure 7. Extended and compact conformation for **M2R1**.

3. QM/MM study of hydrolysis reaction of O-acetyl-propranolol (M0)

As shown in Figure 2, **M0R2d** adopts binding mode IIb and other four conformations are in binding mode Ia. **M0R1b** and **M0R3c** differ in the orientation of the naphthoxy group in the pocket. **M0R1b** and **M0R1c** have more prominent differences in intramolecular hydrogen bonding pattern. **M0R1b'** has similar substrate conformation with **M0R1b**, but has a more open protein conformation.

M0S2d and **M0S3d** adopt binding mode Ib and Ia, respectively (which differ in the orientation of the acyl group in the binding pocket). **M0S1c** adopts binding mode IIb.
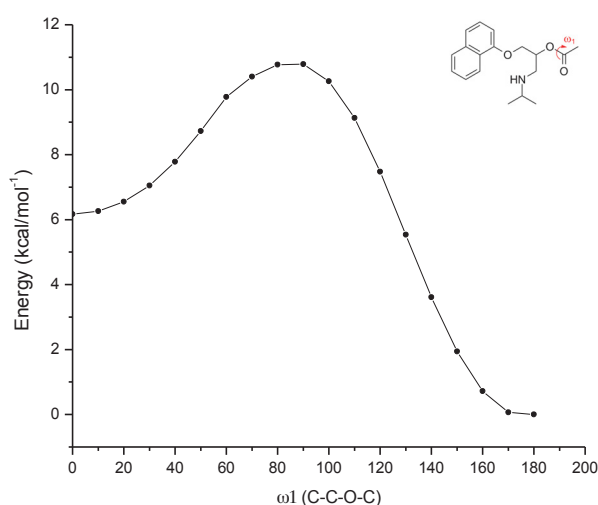
Figure 8. Calculated PES of the rotation of the ester group of M0.

Table 4. QM(B3LYP/TZVP) and MM(CHARMM) energies in kcal/mol (relative to the MCC) for the transformation of **R-M0** for the stationary points in binding mode I and II.

Binding mode I	M0R1b		M0R1c		M0R1b'		M0R3c	
	E_{QM}	E_{MM}	E_{QM}	E_{QM}	E_{MM}	E_{MM}	E_{QM}	E_{MM}
MCC	0	0	0	0	0	0	0	0
TS1	13.0	-0.8	11.5	-0.2	9.1	14.3	15.0	21.6
TI	10.0	0.5	6.4	1.5	5.4	14.6	11.6	22.5
TS2	18.4	6.5	20.0	2.8	16.9	14.7	19.7	20.6
PDC	6.5	11.9	12.2	6.0	9.5	-0.2	7.1	4.5

Binding mode II	M0R2d	
	E_{QM}	E_{MM}
MCC	0	0
TS1	8.3	5.5
TI	6.8	6.1
TS2	14.0	4.7
PDC	-3.2	1.7

Table 5. Essential distances relating with the CH-π interaction between the substrate and the surrounding residues for **M0R1b'** and **M0R3c**. Distances are given in Å.

		MCC	TS1	TI	TS2	PDC
M0R1b'	Ile189:C···Naphthoxy group	4.51	5.01	4.87	4.59	4.21
	M0:C1'···Trp104	3.30	3.55	3.61	3.53	3.53
M0R3c	Ile189:C···Naphthoxy group	3.88	5.03	4.97	4.54	3.87
	M0:C1'···Trp104	3.51	3.66	3.72	3.71	3.61

Table 6. Energy of the QM region in absence of the MM region (QM gas-phase energy, QM_{GP}), and electrostatic contribution of the MM region (QM_{elec} energy) at the stationary points along the QM/MM reaction profiles computed for **R-M0**, relative to the respective MCCs. Energies are presented in kcal/mol.

Stationary point	M0R1b		M0R1c		M0R1b'		M0R3c		M0R2d	
	QM_{GP}	QM_{elec}	QM_{GP}	QM_{elec}	QM_{GP}	QM_{elec}	QM_{GP}	QM_{elec}	QM_{GP}	QM_{elec}
MCC	0	0	0	0	0	0	0	0	0	0
TS1	20.6	-7.6	18.7	-7.3	21.7	-12.5	23.8	-8.8	20.3	-11.9
TI	22.2	-12.2	21.4	-15.0	27.5	-22.0	25.7	-14.1	25.2	-18.4
TS2	26.0	-7.5	27.2	-7.2	27.4	-10.5	30.0	-10.3	22.7	-8.6
PDC	9.7	-3.2	14.0	-1.7	4.5	5.0	3.5	3.6	-6.4	3.2

Table 7. QM(B3LYP/TZVP) and MM(CHARMM) energies in kcal/mol (relative to the MCC) for the transformation of **S-M0** for the stationary points in binding mode I.

Binding mode I	M0S2d		M0S3d		M0S1c	
	E_{QM}	E_{MM}	E_{QM}	E_{MM}	E_{QM}	E_{MM}
MCC	0	0	0	0	0	0
TS1	-	-	-	-	29.0	-8.4
TI	16.7	4.2	14.4	9.6	23.9	-6.4
TS2	22.7	2.2	21.6	12.8	31.3	-5.2
PDC	-8.8	-4.3	-0.3	-3.7	14.1	-2.1

Table 8. Energy of the QM region in absence of the MM region ($QM_{gas-phase}$ energy) and electrostatic contribution of the MM region (QM_{elec} energy) at the stationary points along the QM/MM reaction profiles computed for **M0R2d** and **M0S2d**, relative to the respective MCCs. Energies are presented in kcal/mol.

Stationary point	M0R2d		M0S2d	
	$QM_{gas-phase}$	QM_{elec}	$QM_{gas-phase}$	QM_{elec}
MCC	0	0	0	0
TS1	20.3	-11.9	-	-
TI	25.2	-18.4	23.1	-6.4
TS2	22.7	-8.6	25.4	-2.6
PDC	-6.4	3.2	-8.4	-0.4

Table 9. QM/MM energies, hydrogen bond distance and charges calculated with two different computational methods for **M0R2d**. Distances are given in Å.

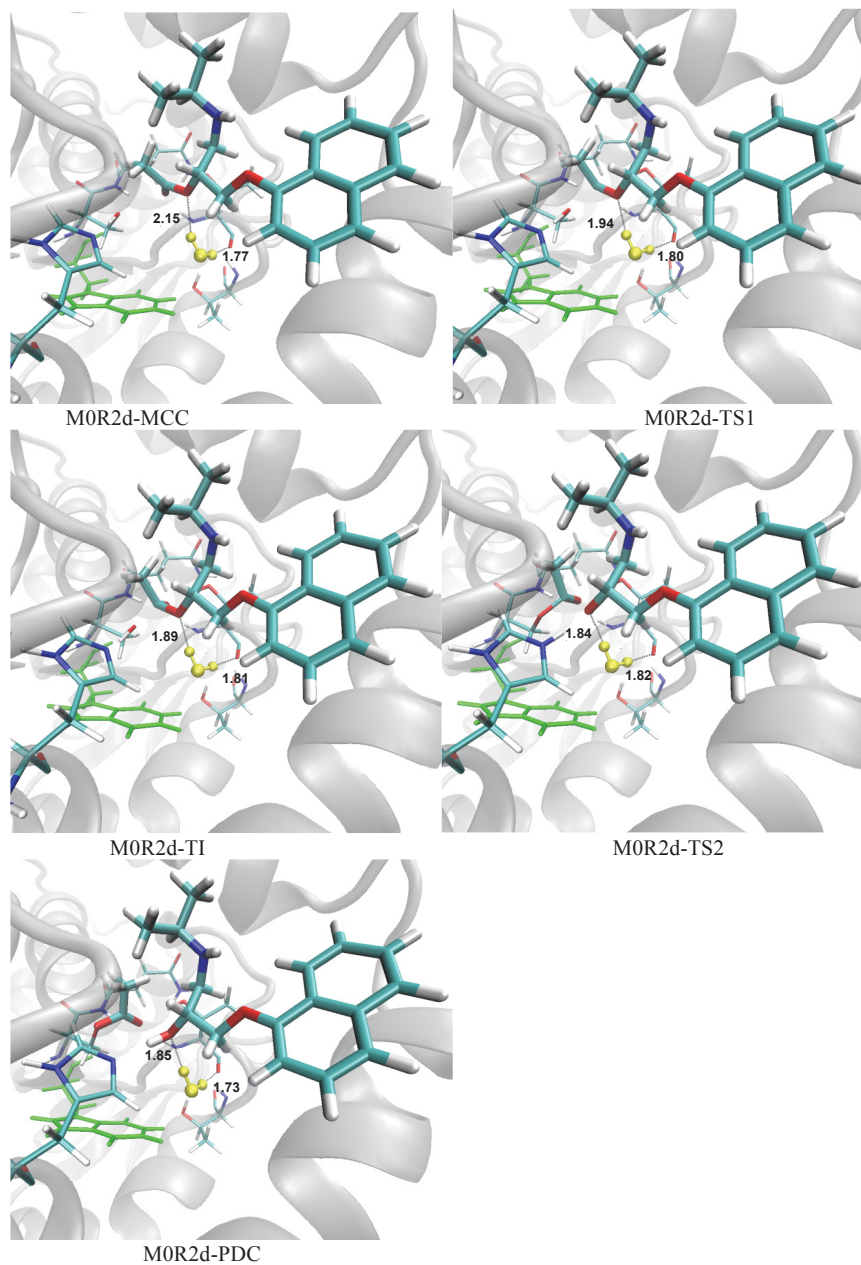| M0R2d		MCC	TS1	TI	TS2	PDC
H₂O in MM region	energy (kcal/mol)	0.0	13.8	12.9	18.7	-1.5
	WAT:H…M0:O2 (Å)	2.15	1.94	1.89	1.84	1.85
	charge: M0:O2	-0.24	-0.27	-0.31	-0.62	-0.49
	charge: WAT:H	0.42	0.42	0.42	0.42	0.42
H₂O in QM region	energy(kcal/mol)	0.0	14.2	13.9	19.8	-1.0
	WAT:H…M0:O2 (Å)	2.75	2.10	2.03	1.95	1.98
	charge: M0:O2	-0.21	-0.27	-0.30	-0.59	-0.47
	charge: WAT:H	0.28	0.28	0.28	0.27	0.27

M0R2d-MCC

M0R2d-TS1

M0R2d-TI

M0R2d-TS2

M0R2d-PDC

Figure 9a. Optimized structures of stationary points for the reaction of **M0R2d**, including a water molecule in the medium pocket, which assists the reaction through hydrogen bond interactions with the ester group.
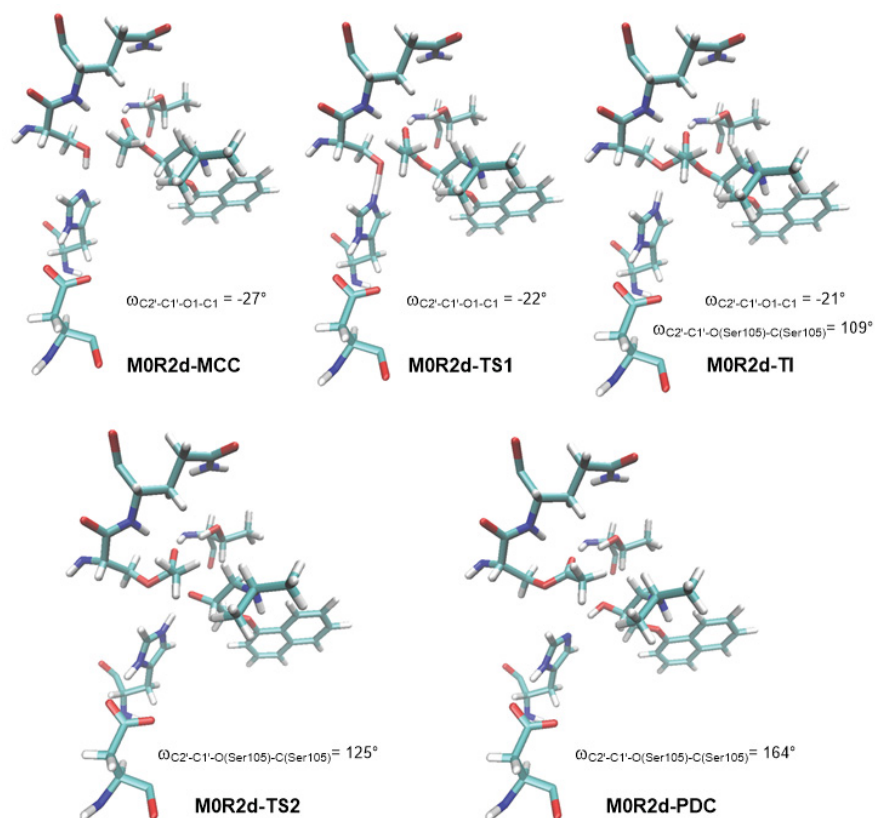
Figure 9b. Optimized structures of stationary points for the reaction of **M0R2d**, including dihedral angels.
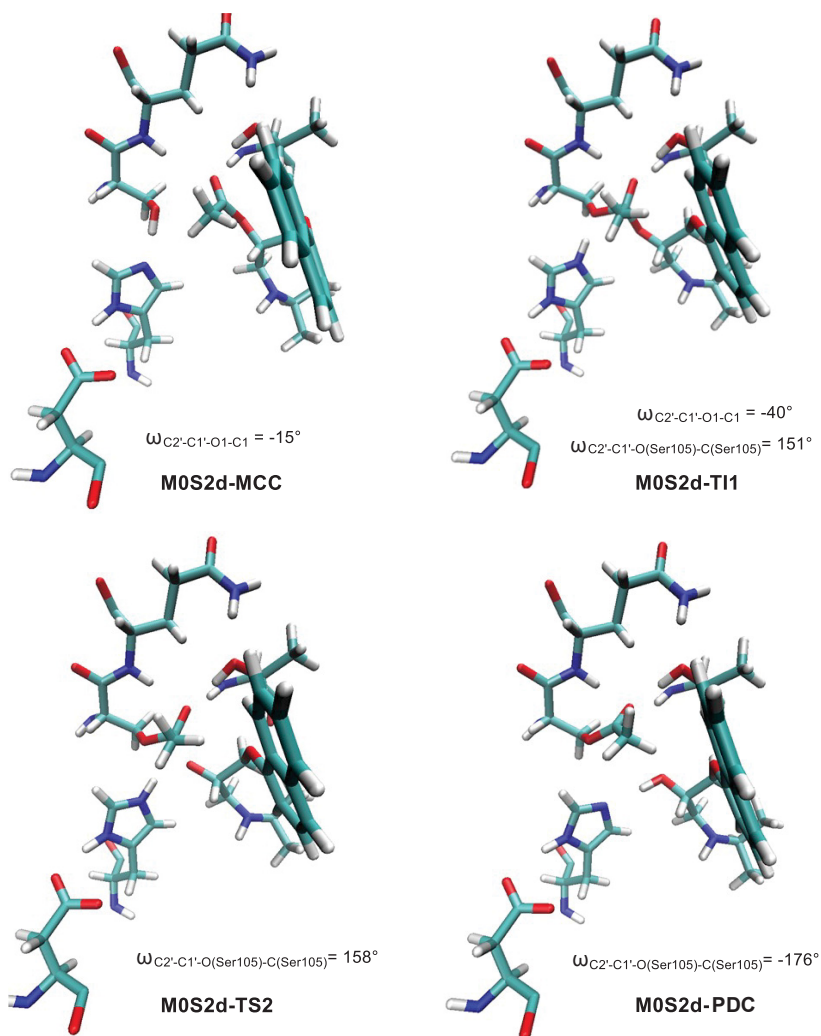
Figure 10. Optimized structures of stationary points for the reaction of **M0S2d**, including dihedral angels.