

# Chemical Science

[www.rsc.org/chemicalscience](http://www.rsc.org/chemicalscience)



ISSN 2041-6539



**EDGE ARTICLE**

D. J. Tantillo, J. B. Siegel *et al.*

Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction

**175** YEARS



Cite this: *Chem. Sci.*, 2016, 7, 4009

# Mechanistically informed predictions of binding modes for carbocation intermediates of a sesquiterpene synthase reaction†

T. E. O'Brien,<sup>a</sup> S. J. Bertolani,<sup>a</sup> D. J. Tantillo<sup>\*a</sup> and J. B. Siegel<sup>\*abc</sup>

Sesquiterpenoids comprise a class of terpenoid natural products with thousands of compounds that are highly diverse in structure, generally containing a polycyclic carbon backbone that is constructed by a sesquiterpene synthase. Decades of experimental and computational studies have demonstrated that these enzymes generate a carbocation in the active site, which undergoes a series of structural rearrangements until a product is formed via deprotonation or nucleophile attack. However, for the vast majority of these enzymes the productive binding orientation of the intermediate carbocations has remained unclear. In this work, a method that combines quantum mechanics and computational docking is used to generate an all-atom model of every putative intermediate formed in the context of the enzyme active site for tobacco *epi*-aristolochene synthase (TEAS). This method identifies a single pathway that links the first intermediate to the last, enabling us to propose the first high-resolution model for the reaction intermediates in the active site of TEAS, and providing testable predictions.

Received 10th February 2016

Accepted 18th March 2016

DOI: 10.1039/c6sc00635c

[www.rsc.org/chemicalscience](http://www.rsc.org/chemicalscience)

## Introduction

Terpenoids comprise one of the largest families of natural products.<sup>1</sup> These metabolites play key roles in the realm of chemical ecology<sup>2</sup> including chemical communication<sup>3,4</sup>—*e.g.*, attracting pollinators,<sup>5</sup> deterring herbivores<sup>6</sup>—heat protection,<sup>7</sup> and chemical warfare.<sup>8</sup> This group of structurally complex metabolites also serves as a diverse chemical library that humans have co-opted for applications in a wide range of areas including agriculture,<sup>9,10</sup> medicine,<sup>11,12</sup> and the development of flavors and fragrances.<sup>13</sup>

Terpenes are constructed in nature from simple precursors by a class of enzymes called terpene synthases (or cyclases). These remarkable enzymes generally transform acyclic, achiral substrates into stereodense, polycyclic products. A mechanism by which many of these enzymes work involves magnesium-assisted ionization of an allylic diphosphate to form an allylic carbocation, which undergoes a combination of intramolecular nucleophilic attacks, hydride and alkyl shifts, and proton transfers.<sup>1,14–20</sup> The enzyme is also thought to play a role in pre-organizing the

conformation of the substrate such that once the carbocation has formed there are a limited number of specific products that will be produced.<sup>1</sup> The results of quantum mechanics (QM) calculations provide support for mechanisms dictated in large part by intrinsic carbocation reactivity.<sup>19,21</sup>

Although impressive strides have been made in recent years,<sup>22–25</sup> the ability to build molecular models of entire carbocation cyclization/rearrangement pathways within an active site has remained elusive for the vast majority of terpene synthases (Fig. 1A). As stated by Major and co-workers in a recent review, “A crucial question in any study of terpene synthases is that of the correct binding mode... indeed, crystal structures of terpene synthases often contain substrates bound in unreactive conformations, partly due to the stickiness of the hydrocarbon moiety of the substrate and its lack of hydrogen bond potential. Thus, there is often great uncertainty regarding the correct binding mode when commencing multi-scale simulation projects of terpene cyclases.”<sup>26</sup> This challenge is likely a major reason why the number of crystal structures available in the protein data bank (PDB) for terpene synthases exceeds 100, while the number of published studies on this family of enzymes that employ QM/molecular mechanics (MM) techniques is less than 10.<sup>19,22,23,26,27</sup>

To tackle this problem we employed a combination of QM calculations to delineate the inherent reactivity of intermediates and computational docking of these intermediates within the enzyme active site. A similar approach was described recently by Jacobson and co-workers in the context of predicting product specificities of terpene synthases;<sup>24,25</sup> however, our focus is on obtaining mechanistic understanding, guiding future mechanistic

<sup>a</sup>Department of Chemistry, University of California Davis, Davis, California, USA.  
E-mail: [djtantillo@ucdavis.edu](mailto:djtantillo@ucdavis.edu); [jbsiegel@ucdavis.edu](mailto:jbsiegel@ucdavis.edu)

<sup>b</sup>Department of Biochemistry and Molecular Medicine, University of California Davis, Davis, California, USA

<sup>c</sup>Genome Center, University of California Davis, Davis, California, USA

† Electronic supplementary information (ESI) available: Additional details on computational methods and results, structures identified in Fig. 2, representative low energy structure for each intermediate docked into 5EAT, files used in the Rosetta docking for each different catalytic motif. See DOI: 10.1039/c6sc00635c

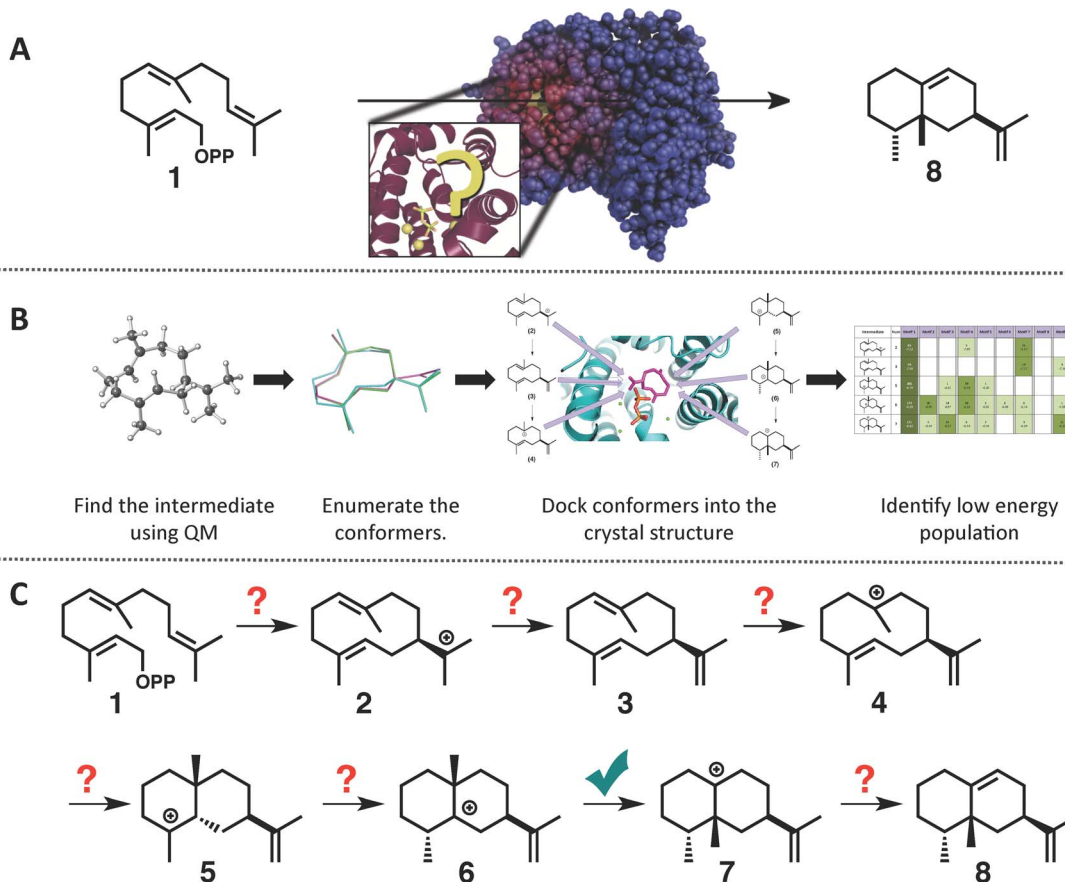


Fig. 1 (A) Cartoon emphasizing the difficulty associated with predicting how the substrate orients in the active site of a terpene synthase. (B) Representation of the workflow employed herein: (1) find the intrinsic energy path with QM, (2) enumerate conformers of the intermediates found, but only keep thermally accessible conformers ( $\leq 5$  kcal mol<sup>-1</sup> higher than the lowest energy conformer), (3) dock the conformers into the crystal structure using Rosetta and constraints based on known chemistry, (4) filter the data to identify in which orientation the low energy population resides. (C) The generally accepted mechanism for the formation of *epi*-aristolochene (**8**) in TEAS; the checkmark indicates the only step for which QM calculations have been reported previously.<sup>37</sup>

experiments and eventually, rational reengineering of this family of enzymes. Here, we employ density functional theory (DFT) calculations rather than semi-empirical calculations,<sup>19</sup> which will provide a more accurate picture of the potential energy landscape over semi-empirical methods, and we take advantage of the unique features of the Rosetta modelling suite<sup>28–30</sup> for docking with the incorporation of experimental and mechanism-based constraints (Fig. 1B).<sup>28,30</sup> In this study we illustrate the utility of this approach for rapidly generating an all-atom model of the reaction pathway for tobacco *epi*-aristolochene synthase (TEAS), a terpene synthase that has been characterized both biochemically and structurally in great detail.<sup>31–34</sup> In a previous study on *epi*-isozizaene synthase,<sup>35</sup> we employed an alternative docking approach that lacks many of the features available in Rosetta.<sup>36</sup> Specifically, Rosetta has the ability to sample significant side chain and backbone conformations, as well as rigid body movements of the ligand. In addition, Rosetta allows the incorporation of user-defined constraints that can encode chemical information about the reaction mechanism that would normally be absent from force-field based molecular modeling approaches.

## Methods

### Quantum mechanics calculations

QM calculations were performed with Gaussian09.<sup>46</sup> Minima and transition state structures (TSSs) were located using mPW1PW91/6-31+G(d,p)<sup>40</sup> with the SMD continuum solvation approach using dichloromethane, chosen as a crude approximation of the mostly nonpolar active site of the enzyme.<sup>47</sup> Stationary points were confirmed as minima or TSSs using harmonic vibrational analysis (no imaginary frequencies for minima; a single imaginary frequency for TSSs). Intrinsic reaction coordinate (IRC) calculations<sup>48</sup> were used to confirm the linkage of a TSS to its flanking intermediates. Fig. 2 shows results for a single set of conformations. All identified intermediates were then subjected to a conformational search using molecular mechanics (MMFF) with Spartan 10.<sup>44</sup> From these conformational searches greater than one hundred structures per intermediate were identified and then fully optimized using Gaussian09 at the mPW1PW91/6-31G(d) level of theory,<sup>40</sup> to minimize computational expense while evaluating over 900 conformers total.<sup>49</sup> All structures found to be within 5 kcal



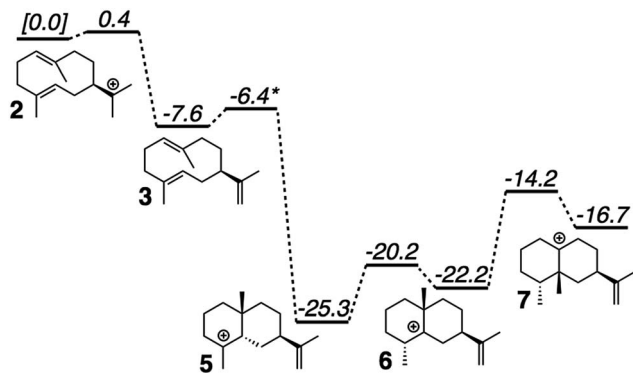


Fig. 2 Computed (mPW1PW91/6-31+G(d,p))<sup>40</sup> relative energies for intermediates and transition state structures involved in the formation of *epi*-aristolochene. \*Best estimate based on a variety of computational experiments (see ESI† for details). Relative energies were put onto the same scale by adding in energies of diphosphate and Tyr/Asp theozyme models (not shown; see ESI† for details).

mol<sup>-1</sup> of the lowest energy conformer were combined to form a library of energetically accessible conformers to be docked into the enzyme.

### Docking calculations

Two crystal structures of TEAS (PDB ID: 5EAT & 4DI5) were minimized using a constrained FastRelax<sup>49</sup> procedure from the Rosetta modeling suite.<sup>28–30</sup> The diphosphate/magnesium complex was extracted from another TEAS crystal structure (PDB ID: 3LZ9)<sup>34</sup>, which was docked (unchanged) along with previously generated conformer libraries into both relaxed crystal structures using the chemically meaningful constraints described below (see ESI† for additional details on constraints as well as the 4DI5 docking results). 2500 docking runs per catalytic orientation per intermediate were carried out to ensure that sampling was sufficient. The resulting structures were combined and then filtered by: (1) their ability to meet the constraints – structures that did not satisfy the constraints were not considered, (2) total protein energy – only structures that were one standard deviation or lower from the mean were considered, (3) interface energy – the top 10% in interface energy were selected from the structures that were in the low total protein energy population. These final filtered structures were then grouped by binding orientation to identify where the low energy population resided (Fig. 4).

### RMSD calculations

The entire protein structures were aligned using TMalign<sup>50</sup> and the RMSD calculation was then performed on each carbon in the skeleton between intermediates, except for the 6 to 7 transition where the shifting methyl group was not considered.

## Results and discussion

### Quantum mechanical modeling of the reaction pathway

The generally accepted mechanism for *epi*-aristolochene formation (Fig. 1C) involves removal of the diphosphate and

subsequent nucleophilic attack on the resulting carbocation to form intermediate 2, deprotonation to form germacrene A (3), reprotonation to form 4, cation–alkene cyclization to form 5, 1,2-hydride shift to form 6, 1,2-methyl shift to form 7, and deprotonation to generate the final product, 8. Previous studies on the energetic viability of this pathway have only examined a small portion of the reaction coordinate (Fig. 1C, checkmark).<sup>37</sup> Therefore DFT calculations (see Methods section for details) were carried out on the pathway shown in Fig. 1C in order to characterize all carbocation intermediates involved in *epi*-aristolochene formation, as well as transition state structures (TSS) connecting them, in terms of both structure and relative energy (in the absence of TEAS). To simulate key portions of the enzyme's active site's effect on the mechanism, two different fragments of the active site were modeled in the QM calculations, termed theozymes. These theozyme calculations have provided insight in the past with defining the role that a key residue might have on the potential energy landscape.<sup>38</sup> In attempts at finding the 2 to 3 TSS a diphosphate (PPI) was used as the base to deprotonate 2. In attempts to find the 3 to 5 TSS a phenol (to represent tyrosine) and acetate (to represent an activating aspartate) were used.

Minima corresponding to each putative intermediate were found with the exception of carbocation 4 (Fig. 2). Geometry optimizations for intermediate 4 consistently generated intermediate 5, suggesting that protonation and cyclization might be concerted.<sup>19</sup> Attempts to locate a TSS that directly linked 3 to 5 in the presence of active site groups (see ESI† for details) led only to TSSs for irrelevant processes, likely a result of a flat energy surface near carbocation 3, *i.e.*, the barrier for conversion of 3 to 5 is very small. In order to evaluate this hypothesis, we performed a scan in which the C–H distance (corresponding to protonation of the C=C  $\pi$ -bond) and C–C distance (for the forming C–C bond) were varied while the remainder of the structure was allowed to relax, in the presence of a theozyme<sup>38</sup> corresponding to the active site Tyr/Asp pair thought to be involved in protonation.<sup>31</sup> The results of this computational experiment allowed us to estimate that the barrier for conversion of 3 to 5 is likely  $\leq 1$  kcal mol<sup>-1</sup> (see ESI† for details). The remainder of the *epi*-aristolochene pathway did not hold any surprises. Conversion of 5 to 6 and 6 to 7 are predicted to be endothermic steps, however, suggesting that an equilibrium between these species may exist in the active site until site-selective deprotonation (*vide infra*) occurs.<sup>35,39</sup>

### Modeling the reaction pathway in TEAS

We then set out to determine the most probable orientations of the carbocation intermediates in the active site of TEAS, in pursuit of an atomic-resolution model of the entire enzyme promoted pathway. The binding mode of the diphosphate is well-defined in the available crystal structures of TEAS.<sup>41</sup> Using these crystal structures we performed molecular docking of each intermediate in the reaction pathway within the TEAS active site. During the docking simulations we employed the use of functional constraints based on mechanistic data to enforce molecular interactions that would be required for the enzyme to



perform the requisite cyclization chemistry leading to the final product, a key advantage of the Rosetta molecular modeling suite. In these simulations we used three explicit constraints involving: (1) the departing diphosphate oxygen that results in carbocation formation; (2) deprotonation of 2; and (3) protonation of 3.

Based on the TEAS crystal structures, three diphosphate oxygen atoms are pointed into the active site. However, it is not known which of these oxygens was connected to the hydrocarbon portion of the substrate. In addition, it has been hypothesized that one of these three oxygens is the base<sup>18,42</sup> that deprotonates 2 to form germacrene A (3), but which oxygen is also unknown. Understanding the relative positions of the components of the ion pair is critical to identifying the most likely orientation of the substrate in the pocket. Recent work has demonstrated the importance of careful positioning of the cation and anion in synthetic terpene-forming cyclization/rearrangement reactions.<sup>43</sup> Given the three possible sites of carbocation disconnection and the three possible sites for deprotonation, nine different possible ion pair orientations arise (Fig. 3, blue and yellow spheres, respectively).

The other mechanism-based constraint was based on the observation reported by Rising *et al.* that mutation of tyrosine 520 to phenylalanine in TEAS changed its activity to that of a germacrene synthase,<sup>31</sup> implying that this tyrosine may play the role of acid, protonating 3 to give 5. Consequently, the phenolic oxygen of Tyr520 was constrained in our docking calculations to reside near to the carbon of the C=C  $\pi$ -bond that is protonated (Fig. 3, red; see Methods section for details).

Finally, before docking we generated conformational libraries for each of the intermediates along the reaction

pathway (Fig. 1B). This was accomplished by subjecting each minimum to a force field-based conformational search followed by a geometry optimization using DFT (at the mPW1PW91/6-31G(d) level of theory) to obtain estimates of each conformation's energy (see Methods section for details).<sup>40,44</sup> Libraries consisting of conformers within 5 kcal mol<sup>-1</sup> of the lowest energy conformer for each intermediate were then docked into crystal structures of TEAS using the Rosetta modeling suite.<sup>28–30</sup>

Twenty five hundred independent docking simulations were run for each intermediate conformer library and each set of constraints (*i.e.*, each possible reaction mechanism). This totaled 112 500 independent docking simulations in order to sample the entire reaction pathway space within TEAS. For each intermediate the solution set was filtered based on satisfaction of the constraints, system energy, and interface energy (see Methods for details). All simulations were done in parallel on two different TEAS crystal structures (5EAT and 4DI5) to ensure subtle structural changes did not dramatically change the docking results and, as expected, there were only minor changes between the results of docking simulations when run using different X-ray structures (see ESI†).

The simulations show a striking enrichment of low energy models for the first reaction mechanism motif with every intermediate (Fig. 4). This result indicates that there is only one binding orientation that links all intermediates to each other – Motif 1. That orientation corresponds to a scenario where the leaving oxygen A (Fig. 3) is also the base that deprotonates intermediate 2. A similar scenario—where the leaving oxygen is also the nucleophile that leads to a diphosphorylated product—was supported by <sup>18</sup>O-labeling experiments for bornyl diphosphate synthase.<sup>45</sup> Interestingly, the orientation of substrate analogs in TEAS active sites in crystal structures (PDB codes: 3LZ9, 3M01, 5EAU)<sup>34,41</sup> would be consistent with Motif 2 (see Fig. F in the ESI†). However, during our simulations very few of the intermediates resulted in low energy conformations in that orientation while satisfying the constraints. This result is consistent with the observation that crystal structures of these enzymes can be misleading in terms of providing insight into the catalytically productive binding mode, especially when conclusions are drawn based on structures with open active sites.<sup>26</sup>

### Structural analysis of reaction pathway from docking

Since a large amount of structural space was sampled during the docking simulations, it is possible that the low energy solutions for any one intermediate are not structurally similar to the low energy solutions for next intermediate. Given how reactive carbocations are (*e.g.*, Fig. 2), we hypothesized that the movement between any two minima is likely to be minimal (*i.e.*, while vibration is reasonable, translation and rotation are less so). Therefore, to obtain an explicit all-atom model of the reaction pathway from beginning to end we identified the low energy structures within Motif 1 with minimal movement between intermediates.

A pair-wise RMSD of carbon atoms was calculated for each pair of connected intermediates for every low energy structure

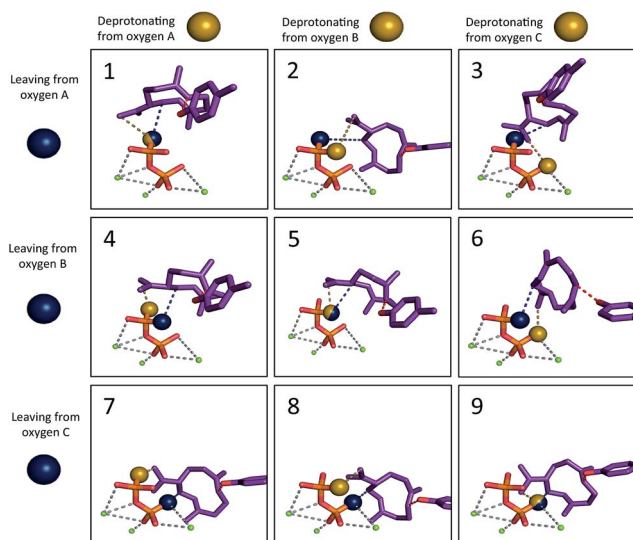


Fig. 3 The nine different ion pair orientations (Motifs 1–9). The blue sphere represents the oxygen that was previously bound to the carbon skeleton and was detached to form the carbocation. The yellow sphere represents the oxygen that would play the role of the base in formation of 3. The tyrosine oxygen–alkene carbon constraint is depicted in red. The grey dashed lines illustrate the coordination of magnesium ions (green) to diphosphate.





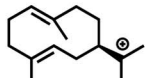
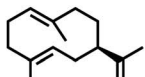
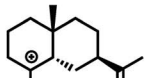
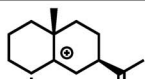
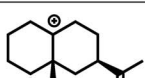
Intermediate	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6	Motif 7	Motif 8	Motif 9
2 	<b>68.4%</b> -7.13			<b>5.3%</b> -7.09			<b>26.3%</b> -6.97		
3 	<b>72.8%</b> -7.59						<b>23.4%</b> -7.11		<b>3.7%</b> -7.16
5 	<b>83.0%</b> -6.74		<b>0.4%</b> -6.42	<b>16.2%</b> -6.54	<b>0.4%</b> -6.36				
6 	<b>67.6%</b> -6.38	<b>7.8%</b> -6.09	<b>7.0%</b> -6.07	<b>11.7%</b> -6.26	<b>0.8%</b> -6.02	<b>2.3%</b> -6.06	<b>2.3%</b> -6.14		<b>0.4%</b> -6.08
7 	<b>71.8%</b> -6.63	<b>2.5%</b> -6.54	<b>10.1%</b> -6.37	<b>2.5%</b> -6.54	<b>2.9%</b> -6.54		<b>1.3%</b> -6.39		<b>8.8%</b> -6.23

Fig. 4 Docking results. Each intermediate is pictured on the left. The darker the green color in each box, the higher the percentage of low energy structures that are found in that catalytic orientation. If no low energy solutions were found for a particular intermediate then no value is given. The number in bold is the percent of total low energy structures found for that that catalytic motif when docking a particular intermediate. The number in italics is the average interface energy (Rosetta energy units) for those structures.

(e.g., carbons in intermediate 2 to 3, 3 to 5, 5 to 6, 6 to 7; Fig. 5). The mean RMSD for the first pair of connected intermediates was 0.88 Å, which indicates that the majority of the structures are in similar structural orientations in the active site. The mean RMSD for each pair increased after the initial 2 to 3 transition. For the 3 to 5 transition the mean RMSD is 2.16 Å, for the 5 to 6 transition it is 1.91 Å and for the 6 to 7 transition it is 2.39 Å. The increased RMSD values for all the transitions after the 2 to 3 transition are due, in part, to a reduction of the ligand volume (see the 3 to 5 overlay in Fig. 6), but are also a result of alternate docking modes.

This is clearest in the 6 to 7 transition, where the average RMSD score appears to have a tri-modal distribution, with one population around 1 Å RMSD, one at 3 Å RMSD and the third at 4.5 Å RMSD. These three populations correspond to three distinct docking orientations (see Fig. G in the ESI†), only one of which aligns well with one of the many docking orientations for intermediate 6. The 2 to 3 transition also appears to have bi-modal character, but in this instance the macrocyclic portion of the structures align very well and the population difference results from two different docking orientations of the exocyclic isopropylene tail in intermediate 3 (see Fig. H in the ESI†), whereas intermediate 2 only has a single docking orientation. If

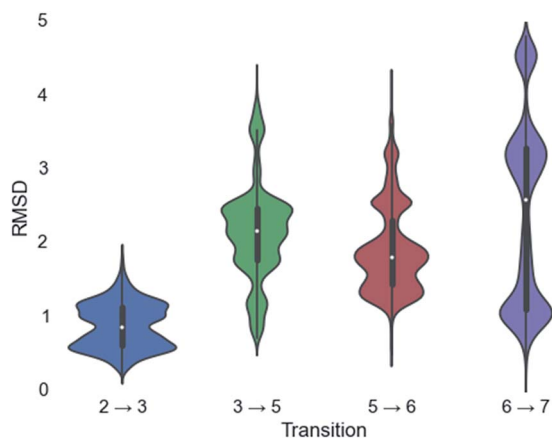


Fig. 5 Violin plots of the RMSD calculations for each transition. The white dot represents the mean, the thicker black region represents the first quartile, and the thin black line represents the standard deviation. Mirrored on both sides of the standard deviation line is the population of any given RMSD score.

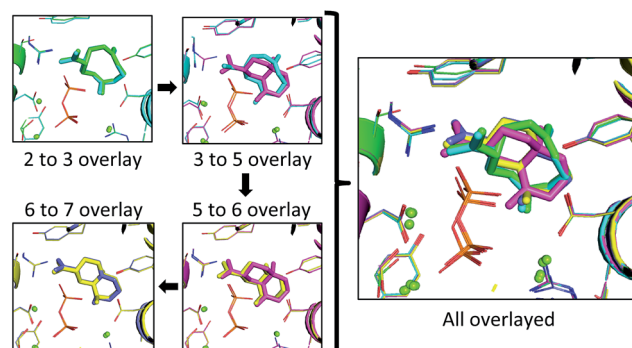


Fig. 6 The lowest RMSD structures for each transition (left) and all lowest RMSD intermediates overlaid (right). Each intermediate and the resulting protein structure is represented by a different color: 2 is represented in green, 3 in cyan, 5 in pink, 6 in yellow and 7 in blue. The green spheres correspond to positions of magnesium ions and the diphosphate is shown as sticks in orange (phosphorus) and red (oxygen).



the RMSD is recalculated without taking the tail into account, the average RMSD score for the 2 to 3 transition decreases to 0.68 Å and a more typical Gaussian distribution is observed (see Fig. 1 in the ESI†).

In the 2 to 3 overlay the overlap is the strongest, with the lowest RMSD for that pair of 0.21 Å, and structures 2 and 3 occupying almost identical spaces (Fig. 6). The 3 to 5 overlay has a higher lowest RMSD value of 0.68 Å, due, in part, to the contraction of volume associated with cyclization (Fig. 6). The lowest RMSD of the 5 to 6 transition is 0.37 Å, which is lower than that for the previous transition. For the 6 to 7 overlay the shifting methyl group was not taken into account for the RMSD calculation, as it would be impossible for it to be in the same position in both structures in that it changes connectivity during the transition. The lowest RMSD for that transition is 0.25 Å. When the lowest RMSD structures for all transitions are overlaid, there is an unambiguous convergence to a single region in the enzyme (Fig. 6, right).

### Predictions to guide future experiments

The binding model presented here leads to (at least) two testable predictions. First, we predict that protonation by Tyr520 would occur on the *re* face of the  $\pi$ -bond; this could be tested by labeling the substrate (*via* chemical synthesis) or the Tyr (*via* the use of D<sub>2</sub>O as solvent). Second, we predict that the diphosphate is not the final base that deprotonates 7 to give 8, since no oxygen in the diphosphate is found to be closer than 5.2 Å from either hydrogen at the position to be deprotonated. We hypothesize that Tyr520 is the final base, with a distance of 2.9 Å in the complex with 7 and an orientation reasonable to remove the pro-*R* hydrogen. This stereochemical prediction could be tested by deuterium labeling of the substrate.

## Conclusions

We present the first all-atom model of the full reaction pathway for TEAS (Fig. 6). Identifying the relative positions of the anionic diphosphate and carbocations in the active sites of terpene synthases is of vital importance for understanding how these enzymes function and provides a basis for further computational studies, *e.g.*, QM/MM dynamics simulations. The method we employ tackles that issue through a combination of QM and computational docking with Rosetta, which allows us to incorporate previous experimental data into the docking, and results in the identification of a single orientation that links all the intermediates along the *epi*-aristolochene-forming pathway. This model allows us to make predictions about the stereochemistry of protonation of 3 and the stereochemistry of deprotonation of 7 (and the identity of the base responsible). Generating such all-atom models of carbocation cyclization/rearrangement pathways in the context of their accompanying protein hosts will enable future efforts to carry out the rational redesign of reaction specificity for this class of enzymes.

## Acknowledgements

TEO acknowledges Vladimir Yarov-Yarovoy for his feedback on this work. D. J. T. gratefully acknowledges support from the National Science Foundation (CHE-0957416 and CHE030089 for supercomputing resources *via* XSEDE). J. B. S. acknowledges the Sloan Foundation (#BR2014-012), ARPA-E (DE-AR0000429), and UC Davis for their support. T. E. O. was supported by the United States Department of Education (GAANN fellowship), the Alfred P. Sloan Foundation URM PhD program and UC Davis' Bradford Borge fellowship. S. J. B. gratefully acknowledges the Chemistry Graduate Program Fellowship (2014) from UC Davis for support.

## Notes and references

- 1 D. J. McGarvey and R. Croteau, *Plant Cell*, 1995, **7**, 1015–1026.
- 2 T. Eisner and J. Meinwald, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 1.
- 3 C. Wicker-Thomas, *J. Insect Physiol.*, 2007, **53**, 1089–1100.
- 4 M. Umehara, A. Hanada, S. Yoshida, K. Akiyama, T. Arite, N. Takeda-Kamiya, H. Magome, Y. Kamiya, K. Shirasu, K. Yoneyama, J. Kyozuka and S. Yamaguchi, *Nature*, 2008, **455**, 195–200.
- 5 K. J. R. P. Byers, J. P. Vela, F. Peng, J. A. Riffell and H. D. Bradshaw, *Plant J.*, 2014, **80**, 1031–1042.
- 6 A. Kessler and I. T. Baldwin, *Science*, 2001, **291**, 2141.
- 7 T. D. Sharkey, A. E. Wiberley and A. R. Donohue, *Ann. Bot.*, 2008, **101**, 5–18.
- 8 R. J. Grayer and T. Kokubun, *Phytochemistry*, 2001, **56**, 253–263.
- 9 D. W. Gammon, M. A. Brown and J. E. Casida, *Pestic. Biochem. Physiol.*, 1981, **15**, 181–191.
- 10 M. B. Isman, *Annu. Rev. Entomol.*, 2005, **51**, 45–66.
- 11 M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon and A. T. McPhail, *J. Am. Chem. Soc.*, 1971, **93**, 2325–2327.
- 12 D. L. Klayman, *Science*, 1985, **228**, 1049–1055.
- 13 S. C. Roberts, *Nat. Chem. Biol.*, 2007, **3**, 387–395.
- 14 D. E. Cane, *Chem. Rev.*, 1990, **90**, 1089–1103.
- 15 D. E. Cane, in *Compr. Nat. Prod. Chem.*, ed. S. D. B. N. Meth-Cohn, Pergamon, Oxford, 1999, pp. 155–200.
- 16 E. M. Davis and R. Croteau, *Top. Curr. Chem.*, 2000, **209**, 53–95.
- 17 D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412–3442.
- 18 D. W. Christianson, *Curr. Opin. Chem. Biol.*, 2008, **12**, 141–150.
- 19 D. J. Tantillo, *Nat. Prod. Rep.*, 2011, **28**, 1035–1053.
- 20 Y. J. Hong and D. J. Tantillo, *J. Am. Chem. Soc.*, 2015, **137**, 4134–4140.
- 21 L. Zu, M. Xu, M. W. Lodewyk, D. E. Cane, R. J. Peters and D. J. Tantillo, *J. Am. Chem. Soc.*, 2012, **134**, 11369–11371.
- 22 M. Weitman and D. T. Major, *J. Am. Chem. Soc.*, 2010, **132**, 6349–6360.
- 23 D. T. Major and M. Weitman, *J. Am. Chem. Soc.*, 2012, **134**, 19454–19462.
- 24 B.-X. Tian, F. H. Wallrapp, G. L. Holiday, J.-Y. Chow, P. C. Babbitt, C. D. Poulter and M. P. Jacobson, *PLoS Comput. Biol.*, 2014, **10**, e1003874.



- 25 J.-Y. Chow, B.-X. Tian, G. Ramamoorthy, B. S. Hillerich, R. D. Seidel, S. C. Almo, M. P. Jacobson and C. D. Poulter, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 5661–5666.
- 26 D. T. Major, Y. Freud and M. Weitman, *Curr. Opin. Chem. Biol.*, 2014, **21**, 25–33.
- 27 N. Chen, J. Zhou, J. Li, J. Xu and R. Wu, *J. Chem. Theory Comput.*, 2014, **10**, 1109–1120.
- 28 J. Meiler and D. Baker, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 538–548.
- 29 A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, in *Methods in Enzymology*, ed. L. J. Michael and B. Ludwig, Academic Press, 2011, vol. 487, pp. 545–574.
- 30 F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic and D. Baker, *PLoS One*, 2011, **6**, e19230.
- 31 K. A. Rising, C. M. Starks, J. P. Noel and J. Chappell, *J. Am. Chem. Soc.*, 2000, **122**, 1861–1866.
- 32 B. T. Greenhagen, P. E. O'Maille, J. P. Noel and J. Chappell, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 9826–9831.
- 33 P. E. O'Maille, A. Malone, N. Dellas, B. Andes Hess, L. Smentek, I. Sheehan, B. T. Greenhagen, J. Chappell, G. Manning and J. P. Noel, *Nat. Chem. Biol.*, 2008, **4**, 617–623.
- 34 J. P. Noel, N. Dellas, J. A. Faraldos, M. Zhao, B. A. Hess, L. Smentek, R. M. Coates and P. E. O'Maille, *ACS Chem. Biol.*, 2010, **5**, 377–392.
- 35 R. P. Pemberton, K. C. Ho and D. J. Tantillo, *Chem. Sci.*, 2015, **6**, 2347–2353.
- 36 During the submission process, a study employing a different docking approach for a different terpene synthase was reported: P. Schrepfer, A. Buettner, C. Goerner, M. Hertel, J. van Rijn, F. Wallrapp, W. Eisenreich, V. Sieber, R. Kourist and T. Bruck, *Proc. Natl. Acad. Sci. U. S. A.*, DOI: 10.1073/pnas.1519680113.
- 37 B. A. Hess, L. Smentek, J. P. Noel and P. E. O'Maille, *J. Am. Chem. Soc.*, 2011, **133**, 12632–12641.
- 38 D. J. Tantillo, C. Jiangang and K. N. Houk, *Curr. Opin. Chem. Biol.*, 1998, **2**, 743–750.
- 39 Y. J. Hong and D. J. Tantillo, *J. Am. Chem. Soc.*, 2011, **133**, 18249–18256.
- 40 C. Adamo and V. Barone, *J. Phys. Chem.*, 1998, **108**, 664–675.
- 41 C. M. Starks, K. Back, J. Chappell and J. P. Noel, *Science*, 1997, **277**, 1815–1820.
- 42 Y. J. Hong and D. J. Tantillo, *Org. Biomol. Chem.*, 2010, **8**, 4589–4600.
- 43 S. V. Pronin and R. A. Shenvi, *Nat. Chem.*, 2012, **4**, 915–920.
- 44 Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio Jr, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Yeh Lin, T. Van Voorhis, S. Hung Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. Min Rhee, J. Ritchie, E. Rosta, C. David Sherrill, A. C. Simmonett, J. E. Subotnik, H. Lee Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2006, **8**, 3172–3191.
- 45 R. B. Croteau, J. J. Shaskus, B. Renstrom, N. M. Felton, D. E. Cane, A. Saito and C. Chang, *Biochemistry*, 1985, **24**, 7077–7085.
- 46 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *GAUSSIAN09*, 2009.
- 47 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 48 S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu and K. Morokuma, *Int. J. Quantum Chem.*, 2015, **115**, 258–269.
- 49 P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding and D. Baker, *Protein Sci.*, 2014, **23**, 47–55.
- 50 Y. Zhang and J. Skolnick, *Nucleic Acids Res.*, 2005, **33**, 2302–2309.

