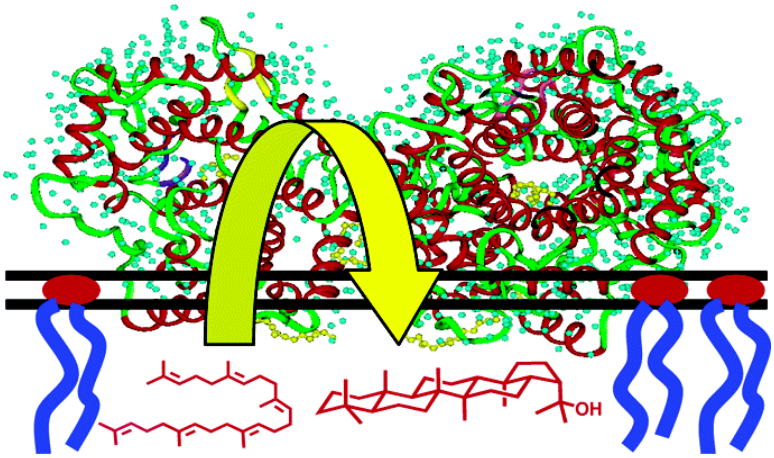
Article

Balancing Kinetic and Thermodynamic Control: the Mechanism of Carbocation Cyclization by Squalene Cyclase

Ramkumar Rajamani, and Jiali Gao

J. Am. Chem. Soc., **2003**, 125 (42), 12768-12781• DOI: 10.1021/ja0371799 • Publication Date (Web): 30 September 2003



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 09/30/2003

Balancing Kinetic and Thermodynamic Control: the Mechanism of Carbocation Cyclization by Squalene Cyclase

Ramkumar Rajamani and Jiali Gao*

Contribution from the Department of Chemistry and Supercomputing Institute, University of Minnesota, Minneapolis, Minnesota 55455

Received July 9, 2003; E-mail: gao@chem.umn.edu

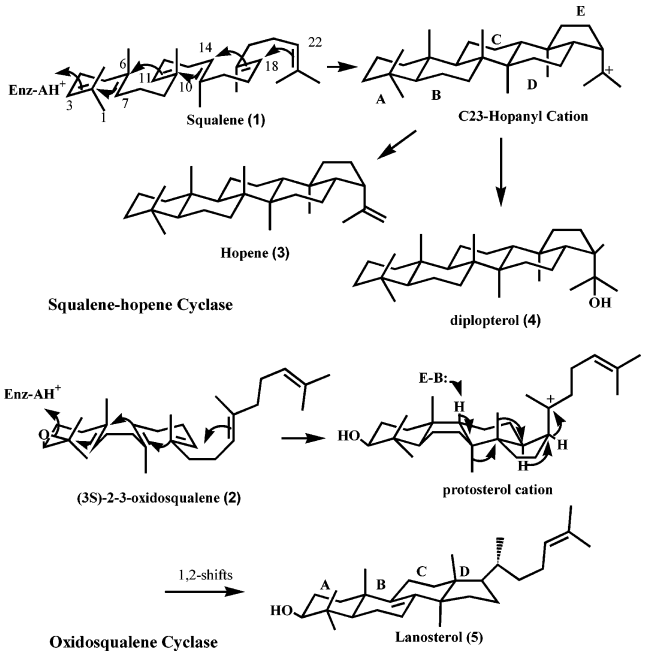
Abstract: Molecular dynamics simulations with a combined quantum mechanical and molecular mechanical (QM/MM) potential have been carried out to investigate the squalene-to-hopene carbocation cyclization mechanism in squalene−hopene cyclase (SHC). The present study is based on free energy simulations by constructing the free energy surface for the cyclization steps along the reaction pathway. The picture that emerges for the carbocation cyclization cascade is a delicate balance of thermodynamic and kinetic control that ultimately favors the formation of the final hopanoids carbon skeleton. A key finding is that the five- to six-membered ring expansion process is not a viable reaction pathway for either C- or D-ring formation in the cyclization reaction. The only significant intermediate is the A/B-bicyclic cyclohexyl cation (**III**), from which two asynchronous concerted reaction pathways lead to, respectively, the 6,6,6,5-tetracyclic carbon skeleton and the 6,6,6,6,5-pentacyclic hopanoids. Experimentally, these two products are observed to have 1% and 99% yields, respectively, in the wild-type enzyme. We conclude that the product distribution in the wild-type enzyme is dictated by kinetic control of these two reaction pathways.

1. Introduction

The remarkable regio- and stereoselectivity of the enzymatic cyclizations of squalene (**1**) and 2,3-oxidosqualene (**2**) to form polycyclic triterpenes have fascinated chemists and biochemists for over half a century.[1−7] These enzymatic reactions yield a variety of triterpenoids, including hopene (**3**) and diplopterol (**4**) (from prokaryotic species), tetrahymanol (from Protozoa), and lanosterol (**5**) (yeast, fungi, and mammals), among other natural products from plants; these triterpenes are precursors of all hopanoids and steroids.[5] The prokaryotic squalene−hopene cyclase (SHC) converts squalene into the pentacyclic hopane skeleton accompanied by the creation of nine stereocenters, all in one step, whereas eukaryotic oxidosqualene cyclases (OSC) catalyze the transformation of 2,3-oxidosqualene into the tetracyclic species, lanosterol, which is further converted into cholesterol (Scheme 1).

Squalene−hopene cyclase catalyzes the cyclization of squalene to form hopene (80%) and diplopterol (20%), which are precursors of a wide variety of hopanoids that condense bacterial membranes.[7,8] SHC is closely related to oxidosqualene cyclase

Scheme 1. Cyclization Reactions by Squalene−Hopene Cyclase and (3*S*)-2,3-Oxidosqualene−Lanosterol Cyclase



in higher organisms; both are membrane-bound with 26% sequence homology that shares a particular sequence similarity with five to eight copies of a highly conserved α-helix−turn motif, called the QW motif,[9,10] which is rich in aromatic residues and unique for this class of enzymes. One long-sought-after

(1) Woodward, R. B.; Bloch, K. *J. Am. Chem. Soc.* **1953**, *75*, 2023−2024.
(2) Cornforth, J. W.; Cornforth, R. H.; Donninger, C.; Popjak, G.; Shimizu, Y.; Ichii, S.; Forchielli, E.; Caspi, E. *J. Am. Chem. Soc.* **1965**, *87*, 3224−3228.
(3) Corey, E. J.; Ortiz de Montellano, P. R. *J. Am. Chem. Soc.* **1967**, *89*, 3362−3363.
(4) Corey, E. J.; Virgil, S. C.; Sarshar, S. *J. Am. Chem. Soc.* **1991**, *113*, 8171−8172.
(5) Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev. (Washington, DC)* **1993**, *93*, 2189−2206.
(6) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 2812−2833.
(7) Hoshino, T.; Sato, T. *Chem. Commun. (Cambridge, U.K.)* **2002**, 291−301.
(8) Seckler, B.; Poralla, K. *Biochim. Biophys. Acta* **1986**, *881*, 356−363.
(9) Tippelt, A.; Jahnke, L.; Poralla, K. *Biochim. Biophys. Acta* **1998**, *1391*, 223−232.

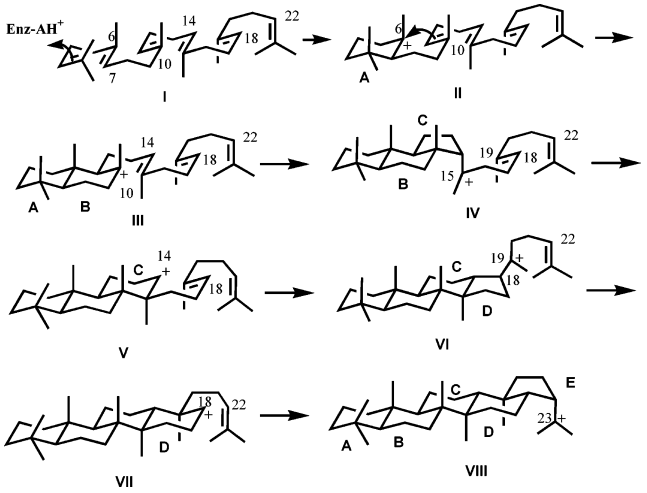
10.1021/ja0371799 CCC: $25.00

question is the detailed mechanism of cyclases, particularly in view that synthetic experimental efforts have failed to duplicate the remarkable one-step transformation of the squalene and 2,3-oxidosqualene cyclization reaction without enzyme catalyst.[5,6]

In OSC, the substrate is enforced to adopt a chair−boat−chair conformation and the protosterol cation intermediate undergoes a series of 1,2-hydride and methyl group shifts to yield lanosterol.[5,6] On the other hand, squalene takes an all-chair conformation in SHC without further carbon skeleton changes after the initial cyclization (Scheme 1). Despite these differences, the catalytic mechanisms of OSC and SHC are believed to be similar on the basis of results from extensive chemical and biochemical studies.[5−7] However, there is still a lack of detailed theoretical investigation of the mechanism of the enzymatic carbocation cyclization.[11,12] The availability of the X-ray structure for squalene−hopene cyclase[13−16] from *Alicyclobacillus acidocaldarius* makes computational studies possible, and this paper reports the results from a combined quantum mechanical and molecular mechanical (QM/MM) free energy simulation of the energetics and reaction pathways for the conversion of squalene to hopene catalyzed by squalene cyclase.

Site-directed mutagenesis experiments coupled with X-ray crystal structure of SHC revealed that Asp376, which is stabilized by His451 with the aid of the neighboring, tightly hydrogen-bonded Asp374−Asp377 pair and Tyr495, is the catalytic acid, and the position of the Asp376 proton has been proposed to be at the anti orientation, which is several pK_a units more acidic than the corresponding syn orientation.[13,14,17−19] In OSC, Corey et al.[20−22] demonstrated experimentally that a similar acid residue is the catalytic acid, and the oxirane opening and A-ring formation are concerted. The latter was also supported by calculations using ab initio molecular orbital and density functional theory.[12] Presumably, the protonation and A-ring formation in SHC is also concerted, especially in view of the relatively nonbasic nature of 2,3-olefin in squalene, which would benefit from anchimeric assistance from the 6,7-π electrons.[6] Unlike the reaction of OSC, chemical studies have not provided mechanistic insights on the sequence of A- and B-ring formation in SHC. Scheme 2 depicts the mechanistic steps for the SHC reaction proposed in the review articles by Wendt et al.[6] and by Hoshino and Sato;[7] throughout this paper, we use Roman numerals to specify carbenium ion intermediates, while the carbon atoms are numbered on the basis of squalene.

Scheme 2. Proposed Mechanism for the Squalene to Hopene Cyclization in Squalene−Hopene Cyclase[a]



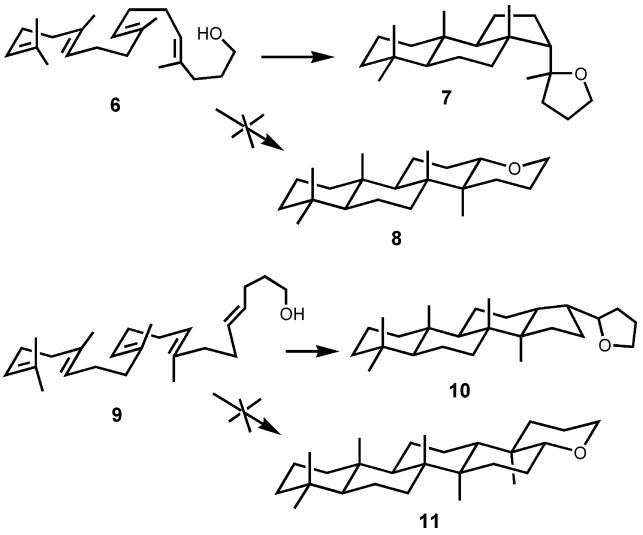
[a] Based on biochemical and X-ray crystallographic experiments.[7] This scheme was used as an initial guide in designing the computational procedure described in this paper.

The formation of both C- and D-rings follows formally two anti-Markovnikov additions, which are thermodynamically disfavored (Scheme 2).[11] The F601A mutation resulted in the isolation of two 6,6,5-fused tricyclic products and one 6,6,6,5-fused tetracyclic compound along with hopene and diplopterol.[23,24] Poralla and co-workers[25] reported that a small amount (ca. 1%) of the products of the wild-type SHC reaction in fact consists of several 6,6,6,5-fused tetracyclic species. These findings led to the proposal that both the six-membered C- and D-rings are, respectively, formed through ring expansions from the thermodynamically favored cyclopentylcarbinyl intermediates, **IV** and **VI**, which implies that C- and D-ring formation are stepwise processes.[6,7,26] The five- to six-membered ring expansion mechanism has an analogy in the OSC reaction.[27] This mechanism also appears to be supported by the observation that the 6,6,5-fused tricyclic product containing a tetrahydrofuran ring (**7**) rather than the 6,6,6,6-fused product (**8**) was obtained from SHC when a truncated substrate (**6**) with a nucleophilic hydroxy group was used (Scheme 3).[24] An all-prechair conformation of the substrate would favor the formation of **8**. The same argument can be made for D-ring formation with the isolation of **10** from the substrate **9**.[24]

Although the stepwise mechanism for C- and D-ring formation would seem to be consistent with site-directed mutagenesis experiments, the 6,6,6-fused tricyclic intermediate, from the C-ring five- to six-membered ring expansion, has never been trapped in mutation studies,[6,7] suggesting that the six-membered C-ring species may not be a discrete intermediate. Furthermore, only trace amounts of two 6,6,6,6-fused tetracyclic products have ever been isolated in all mutation studies,[28] also indicating that

(10) Godzina, S. M.; Lovato, M. A.; Meyer, M. M.; Foster, K. A.; Wilson, W. K.; Gu, W.; De Hostos, E. L.; Matsuda, S. P. T. *Lipids* **2000**, *35*, 249−255.
(11) Jenson, C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1997**, *119*, 10846−10854.
(12) Gao, D.; Pan, Y.-K.; Byun, K.; Gao, J. *J. Am. Chem. Soc.* **1998**, *120*, 4045−4046.
(13) Wendt, K. U.; Poralla, K.; Schulz, G. E. *Science (Washington, DC)* **1997**, *277*, 1811−1815.
(14) Wendt, K. U.; Lenhart, A.; Schulz, G. E. *J. Mol. Biol.* **1999**, *286*, 175−187.
(15) Lenhart, A.; Weihofen, W. A.; Pleschke, A. E. W.; Schulz, G. E. *Chem. Biol.* **2002**, *9*, 639−645.
(16) Lenhart, A.; Reinert, D. J.; Aebi, J. D.; Dehmlow, H.; Morand, O. H.; Schulz, G. E. *J. Med. Chem.* **2003**, *46*, 2083−2092.
(17) Feil, C.; Suessmuth, R.; Jung, G.; Poralla, K. *Eur. J. Biochem.* **1996**, *242*, 51−55.
(18) Sato, T.; Hoshino, T. *Biosci., Biotechnol., Biochem.* **1999**, *63*, 2189−2198.
(19) Gao, J.; Pavelites, J. J. *J. Am. Chem. Soc.* **1992**, *114*, 1912−1914.
(20) Corey, E. J.; Cheng, H.; Baker, C. H.; Matsuda, S. P. T.; Li, D.; Song, X. *J. Am. Chem. Soc.* **1997**, *119*, 1289−1296.
(21) Corey, E. J.; Cheng, H.; Baker, C. H.; Matsuda, S. P. T.; Li, D.; Song, X. *J. Am. Chem. Soc.* **1997**, *119*, 1277−1288.
(22) Corey, E. J.; Stass, D. *J. Am. Chem. Soc.* **1998**, *120*, 3526−3527.
(23) Merkofer, T.; Pale-Grosdemange, C.; Wendt, K. U.; Rohmer, M.; Poralla, K. *Tetrahedron Lett.* **1999**, *40*, 2121−2124.
(24) Hoshino, T.; Kouda, M.; Abe, T.; Ohashi, S. *Biosci., Biotechnol., Biochem.* **1999**, *63*, 2038−2041.
(25) Pale-Grosdemange, C.; Feil, C.; Rohmer, M.; Poralla, K. *Angew. Chem., Int. Ed.* **1998**, *37*, 2237−2240.
(26) Sato, T.; Hoshino, T.; Abe, T. *Chem. Commun. (Cambridge, U.K.)* **1998**, 2617−2618.
(27) Corey, E. J.; Virgil, S. C.; Cheng, H.; Baker, C. H.; Matsuda, S. P. T.; Singh, V.; Sarshar, S. *J. Am. Chem. Soc.* **1995**, *117*, 11819−11820.
(28) Hoshino, T.; Abe, T.; Kouda, M. *Chem. Commun. (Cambridge,U.K.)* **2000**, 441−442.

Scheme 3. Synthetic Substrates That Have Been Used To Trap the 6,6,5-Tricyclic and 6,6,6,5-Tetracyclic Carbenium Ion Intermediates[24]



the six-membered D-ring cation intermediate may not be a stable species in the reaction pathway. This raises an interesting question on whether the isolation of the aborted cyclization products in the mutant enzymes truly provides proof of the reaction pathway in the wild-type enzyme or, rather, they only reflect the altered reaction pathways in the mutants. The observation of aborted cyclization products, in fact, indicates that the corresponding carbocation intermediates must be stabilized so that they existed for an extended period of time to be trapped during the cyclization process. If the carbocation intermediates were destabilized, the reaction rate for the next cyclization step would have been facilitated and the cyclization process would have not been aborted. Therefore, the findings from site-directed mutations may not be evidence for the existence of the corresponding carbenium ion intermediates along the reaction pathway in the wild-type enzyme.

To answer these questions, we need to consider the possibility of a mechanism that involves concerted formation of the six-membered C-ring and electrophilic addition to produce the five-membered D-ring intermediate. Alternatively, another possibility is an extended, concerted mechanism that involves two anti-Markovnikov additions in C- and D-ring formation along with the simultaneous E-ring formation to produce the final hopenoid skeleton. The present study addresses these questions by examining the free energy surface for the carbocation cyclization processes. As a prelude, we find that the enzyme squalene−hopene cyclase uses a delicate balance of thermodynamic and kinetic control in the formation of key carbocation intermediates along the reaction pathway, leading to hopanyl cation as the major final cyclization intermediate. We compare the computed free energy surfaces for the cyclization reactions with experimental findings where byproducts have been trapped in mutants.

In the present study, we do not directly address the issue of the carbocation initialization in the SHC catalysis.[6,12] This will be investigated in a future study, which is concerned with the acidity of the active-site residue Asp374 in SHC and its protonation of the 2,3-olefin in the substrate. Although we will not answer the question regarding the concertedness of the initial protonation and A-ring formation step (**1** → **II**), the present work sheds light on the mechanism of the subsequent cyclization steps

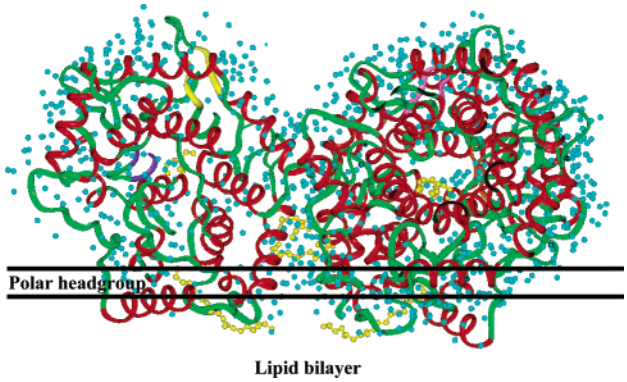


Figure 1. Proposed dimer binding contact with lipid for monotopic membrane protein squalene−hopene cyclase based on the X-ray crystal structure.[13] The active site is located near the center of the α-helix barrel, which is occupied by an inhibitor (see text) shown as a ball-and-stick model in yellow. Several other inhibitor molecules are seen in the pathway at the entrance to the active site and at the interface of the dimer.

important to regio- and stereocontrol in product formation (Scheme 2).

In the following, we present computational details concerning the representation of the protein−substrate system, the validation and accuracy of our computational model, and the dynamics simulation of the free energy surface. Then, we present results and discussion of the reaction mechanism. The paper concludes with a summary of major findings of the present study.

2. Computational Details

2.1. Structural Features. The crystal structure of the bacterial enzyme squalene−hopene cyclase from *Alicyclobacillus acidocaldarius* was first determined by Wendt et al.[13,14] at a resolution of 2.9 Å and later refined to a resolution of 2.0 Å. The medium-resolution structure was obtained in the presence of a competitive inhibitor, *N*,*N*-dimethyldodecylamine *N*-oxide (LDAO),[13] whereas a detergent molecule was found in the active site of the higher resolution structure.[14] In the latter case, the catalytic acid residue Asp376 was also mutated to a cysteine. SHC is a dimeric protein, consisting of 631 amino acid residues in each subunit with a molecular mass of approximately 72 kDa.[13,29] The crystal structure of SHC shows two topological domains: domain 1 is an α6−α6 barrel of two concentric rings of parallel α helices, and domain 2 is an α−α barrel. The active site of SHC occupies a large cavity of 1200 Å^3 between the two domains that is enclosed by the amino ends of the α helices and loops (Figure 1). The active site is accessible through a channel formed by the helices of domain 2, and the channel entrance is surrounded by a hydrophobic plateau of about 1600 Å^2 of accessible area. This observation led Wendt et al.[13] to propose that this plateau is the membrane binding site for this monotopic membrane protein, and the channel provides a path for the substrate to diffuse into the active site and for the product to exit (Figure 1).

Both polar and nonpolar residues are present in the active site of SHC, which is schematically shown in Figure 2. A number of polar residues, considered to be critical to the initiation of the carbocation cyclization process, are found at the top of the active center (Figure 2). The rest of the cavity is composed of a series of aromatic and aliphatic residues, which are thought to play a role in the stabilization of carbocations formed during the cyclization process (through cation−π interactions).[7,13,30−33] However, it should be noted that the

(29) Ochs, D.; Kaletta, C.; Entian, K. D.; Beck-Sickinger, A.; Poralla, K. *J. Bacteriol.* **1992**, *174*, 298−302.
(30) Dougherty, D. A. *Science (Washington, DC)* **1996**, *271*, 163−168.
(31) Ma, J. C.; Dougherty, D. A. *Chem. Rev. (Washington, DC)* **1997**, *97*, 1303−1324.
(32) Gao, J.; Chou, L. W.; Auerbach, A. *Biophys. J.* **1993**, *65*, 43−47.

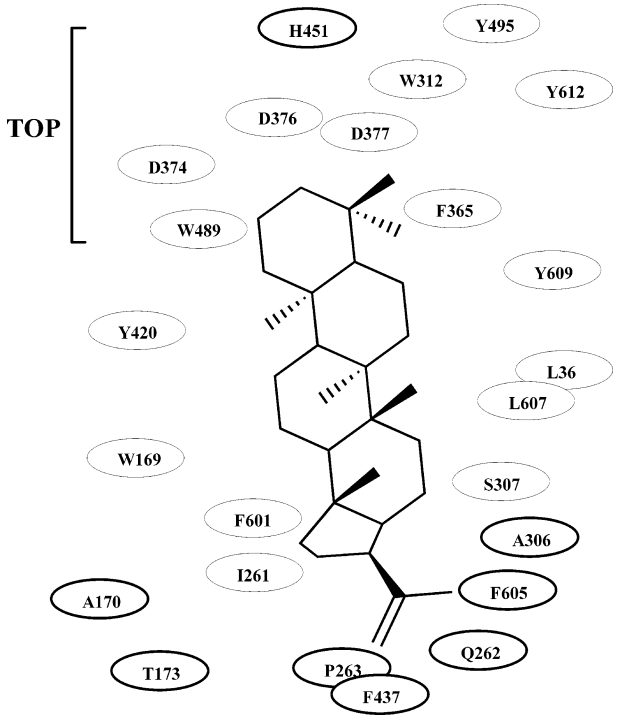


Figure 2. Schematic representation of the relative locations of active-site residues. The hopene−enzyme complex model was optimized by Wendt et al.[13]

formation of the carbocation intermediates in the cyclization reaction is accompanied by the release of a total of about 60 kcal/mol in free energy. So, these intermediates actually need not be especially stabilized at all. Furthermore, if any of the carbocations were stabilized, it would have a greater chance of being trapped or aborted in the cyclization pathway. In fact, it is critical that the energy released in the cyclization process be rapidly dissipated. From our studies, a major role of these aromatic and aliphatic residues seems to provide steric hindrance to ensure proper substrate folding and to protect the carbocations from interacting with any potential nucleophiles. On the basis of the crystal structure, Wendt et al.[13] suggested that the QW motif may be responsible for keeping the protein integrity against a large amount of energy produced during the cyclization process.

2.2. Michaelis Complex Model. An essential and critical step in modeling enzymatic reactions is to construct an initial configuration for the enzyme−substrate system representing the Michaelis complex for the chemical step to be investigated. This typically requires some forms of substrate docking and model-building because the Michaelis complex structure is generally not amenable for X-ray structural determination. In the present study, we used the wild-type enzyme crystal structure determined by Wendt et al.[13] (PDB code 1SQC), which is complexed with an inhibitor in the active site. Previously, Wendt et al. examined the interactions and functional roles of amino acid residues in the active site of SHC. They placed the product hopene, **3**, in the active site with the C_3 atom pointing toward the acidic residue Asp376, and the orientation of **3** in the active site was optimized by rotating the pentacyclic system in 10° increments followed by energy minimization. The coordinates corresponding to rings A, B, and C in this model have been deposited in the protein databank along with the enzyme coordinates.[13] On the basis of this information, we constructed the remaining coordinates for hopene using the builder module in the molecular modeling program InsightII. Our own energy optimization yielded similar results in the orientation of hopene as that found by Wendt et al.[13,14] The resulting conformation was further energy-minimized for 50 steps by the adopted-basis Newton−Raphson (ABNR) method[34] along with the combined QM-AM1/CHARMM potential (see below).[35]

The product structure was then converted into the substrate coordinates of the Michaelis complex. Since the present study is concerned only with the carbocation cyclization process starting from **I**, the C_3-protonated squalene (**I**) is considered to be the initial reactant species in the Michaelis complex. We created a topology file[34] for **I** to be used in the CHARMM force field, separated the carbon−carbon bond distances in **3** that are formed in the cyclization reaction to a value of 3.0 Å, and added a proton at the C_3 carbon with the aid of the H-build module in CHARMM.[34] The resulting substrate configuration was relaxed by carrying out 200 steps of energy minimization by the ABNR method and the CHARMM force field (standard force field for alkenes were used for squalene). This was followed by 50 steps of QM/MM minimization (see below) to obtain a better description of the substrate geometry, with squalene cation **I** treated by the semiempirical AM1 method.[36] In this process, atomic positions of the protein and crystal waters were held fixed. This structure of the SHC−substrate cation complex is used as the starting configuration for the initial equilibration in molecular dynamics simulations.

2.3. Combined QM/MM Potential. A major goal of this study is to determine the two-dimensional potential of mean force for the carbocation cyclization process in squalene cyclase in order to provide insights into the mechanism of the highly controlled cyclization reaction. To this end, we have used a combined quantum mechanical and molecular mechanical (QM/MM) approach,[37−40] in which the substrate, protonated squalene and its subsequent intermediates, is treated quantum mechanically and the rest of the protein−solvent system is represented by the CHARMM22 force field[35] for protein and the three-point-charge TIP3P model for water.[41] The electronic structure of the substrate, which contains 81 atoms, is described at the semiempirical AM1 (Austin model 1) level.[36] There is no covalent bond formed between the substrate and enzyme in the cyclization reaction, and there is no need to introduce a covalent QM/MM boundary.[42] Combined QM/MM methods have been described previously and reviewed in several publications; [38,40,43,44] interested readers are directed to these articles for additional technical details. In this section, we focus on validation of the semiempirical model for reactions involving carbocation intermediates.

It is, of course, desirable to use ab initio molecular orbital theory or density functional theory to treat the substrate, which in general would be accurate and more systematic.[45−50] However, such computations are still too time-consuming to be practical to do free energy simulations for such a large system, particularly in view of the need to generate two-dimensional potentials of mean force. Thus, we adopted the semiempirical AM1 model for the QM region. It perhaps should be mentioned that the development of semiempirical methods has been tightly correlated with the study of the properties of carbocations.[36,51−53]

(33) Mo, Y.; Subramanian, G.; Gao, J.; Ferguson, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 4832−4837.
(34) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(35) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(36) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(37) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(38) Field, M. J.; Bash, P., A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(39) Gao, J.; Xia, X. *Science* **1992**, *258*, 631−635.
(40) Gao, J. In *Rev. Comput. Chem.*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1995; Vol. 7, pp 119−185.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(42) Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714−4721.
(43) Amara, P.; Field, M. J. In *Computational Molecular Biology*; Leszczynski, J., Ed.; Elsevier: Amsterdam, 1999; pp 1−33.
(44) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

Scheme 4. Model Reactions for the Formation of the Tertiary Cyclohexyl Cation from a Tertiary Acyclic Carbenium Ion and for the Five- to Six-Membered Ring Expansion Involving a Cyclopentylcarbinyl Cation to a Secondary Cyclohexyl Cation

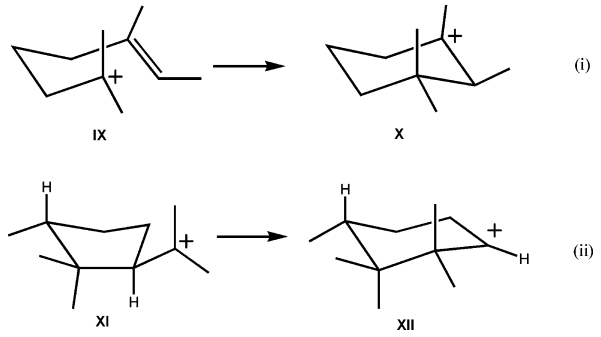


Table 1. Computed Energy of Reaction for Model Reactions i and ii in Scheme 4[a]

reaction	AM1	HF/6-31G(d)	B3LYP/6-31G(d)
i	−13.5	−14.1	−13.7
ii	14.9	13.0	13.5

[a] Model reaction i shows formation of a tertiary cyclohexyl cation; model reaction ii is a five- to six-membered ring expansion. Energy of reaction is computed from ab initio molecular orbital and density functional theory and from semiempirical AM1 calculations. Energies are given in kilocalories per mole.

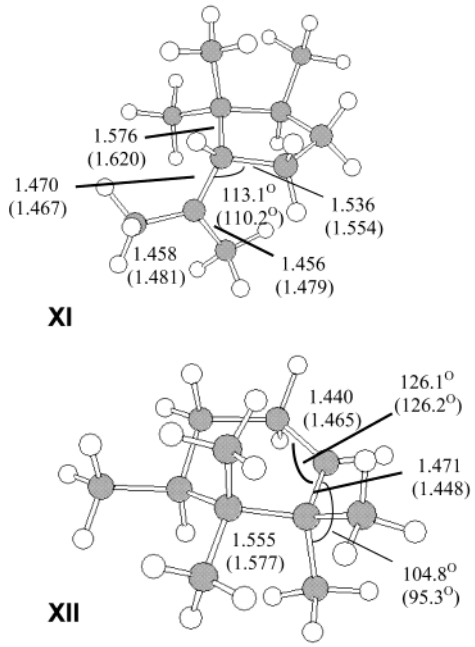


Figure 3. Optimized geometries for the model compounds at the HF/6-31G(d) (values in parentheses) and AM1 levels of theory. Bond distances are given in angstroms and bond angles are in degrees.

It can be anticipated that the semiempirical AM1 method will yield excellent results in comparison with high-level ab initio calculations. Indeed, this expectation has been confirmed by studies of model reactions relevant to the present squalene cyclization reaction.

Two model reactions have been examined, corresponding to (i) the formation of A- and B-ring fragments in **3** and (ii) the five- to six-membered ring expansion in C- and D-ring formation (Scheme 4). For each system, geometry optimizations were carried out at the Hartree−Fock (HF)/6-31G(d),[54] and AM1 levels of theory and energy calculations that include electron correlation were performed with hybrid density functional theory, B3LYP/6-31G(d)//HF/6-31G(d).[55] The results are listed in Table 1. The energy changes for conversion of 2,6-dimethyloct-6-en-2-yl cation **(IX)** to 1,1,2,3-tetramethyl-3-cyclohexyl cation **(X)** are −14.1, −13.7, and −13.5 kcal/mol from HF/6-31G(d)//6-31G(d), B3LYP/6-31G(d)//HF/6-31G(d), and AM1 calculations, respectively. For the ring expansion reaction from 2-(2,2,3-trimethylcyclopentyl)-2-propyl cation **(XI)** to 1,1,2,2,3-pentamethyl-6-cyclohexyl cation **(XII)**, the computed energy changes are, respectively, 13.0, 13.5, and 14.9 kcal/mol at the three levels noted above. This process has also been examined previously by Jenson and Jorgensen[11] at the same levels of ab initio theory, and they noted that an alternative envelope structure of **XI** with the methyl group on the methine carbon of the ring in a pseudoaxial orientation was obtained and it was 0.6 kcal/mol higher in energy than the conformer shown in Figure 3. We have used the more stable conformer for the comparison. These researchers also computed the enthalpy and free energy of the ring expansion reaction, which are 11.9 and 13.4 kcal/mol at the HF and B3LYP level, respectively. Another relevant quantity in their study is the energy of formation for **XI** from the extended, lowest-energy conformation structure, which was predicted to be −9.6 kcal/mol at the HF/6-31G(d)//HF/6-31G(d) level;[11] AM1 calculations yielded a value of −10.7 kcal/mol. These results demonstrate that the semiempirical AM1 model can yield reliable results for carbocation addition and cyclization reactions with an accuracy comparable to the B3LYP/6-31G(d) level.

Figure 3 shows key geometrical parameters determined by HF/6-31G(d) and AM1 calculations for carbocation species **XI** and **XII**. The agreement between ab initio and AM1 results for all model compounds is also good with average errors of 0.04 Å in bond length and 5.5° in bond angle.

2.4. Stochastic Boundary Molecular Dynamics. A stochastic boundary molecular dynamics simulation method was used in the present study.[56−58] Although squalene cyclase is membrane-bound, the active center is located over the lipid interface, primarily in the aqueous region (Figure 1).[13,14] Since the present study is concerned with the chemical steps in the active site, membrane−enzyme interactions will not be explicitly included in the computational model; these interactions are nevertheless an integral part of the overall SHC function, particularly in substrate binding and product release. To mimic the aqueous environment, the enzyme system is surrounded by a 24 Å sphere of water with the center of origin located at the C_{12} position of squalene cation **I**, which is roughly the center-of-mass position. Water molecules that are within 2.5 Å of non-hydrogen atoms in the protein are removed from the system, and all crystal water molecules within this sphere are retained.[59] In SBMD, atoms within 20 Å from the active center are treated by Newtonian molecular dynamics, while the region between

(45) Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J. Phys. Chem.* **1993**, *97*, 11868−11870.
(46) Alhambra, C.; Byun, K.; Gao, J. *ACS Symp. Ser.* **1998**, *712*, 35−49.
(47) Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **1999**, *20*, 1468−1494.
(48) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *110*, 10452−10467.
(49) Rothlisberger, U.; Carloni, P.; Doclo, K.; Parrinello, M. *J. Biol. Inorg. Chem.* **2000**, *5*, 236−250.
(50) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182−3188.
(51) Dewar, M. J. S.; Dieter, K. M. *J. Am. Chem. Soc.* **1986**, *108*, 8075−8086.
(52) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.
(53) Olah, G. A.; Schleyer, P. v. R., Eds. *Carbonium Ions*; Wiley: New York, 1968−1973; Vol. I−IV.
(54) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.
(55) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. E2 ed.; Gaussian, Inc.: Pittsburgh, PA, 1994.
(56) Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6325.
(57) Brunger, A. T.; Brooks, C. L., III; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8458−8462.
(58) Brooks, C. L., III; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159−181.
(59) Alhambra, C.; Gao, J. *J. Comput. Chem.* **2000**, *21*, 1192−1203.

20 and 24 Å is modeled by Langevin dynamics with friction coefficients of 200 ps^{-1} for protein atoms and 62 ps^{-1} for water molecules.[56,58] The rest of the protein system is treated as a reservoir region with the atomic coordinates fixed throughout the simulation.

The solvation procedure was then followed by a short period of equilibration for 10 ps at 100 K with harmonic constraints on nonhydrogen atoms by a force constant of 20 kcal mol^{-1} $Å^{-2}$. Then, the temperature of the system was gradually brought up from 100 K to a final value of 298 K in a heating increment of 1 K/50 steps of molecular dynamics simulation. Next, the harmonic restraining forces were gradually reduced to zero in 25 ps, and an additional 25 ps of unconstrained SBMD simulation was performed. After the initial heating and equilibration step, the water droplet was again overlaid over the active center of SHC, and the procedure described above was repeated. This resulted in a total of 13 109 atoms in the system. Finally, the potential energy surface was switched to the combined QM/MM (AM1/CHARMM-TIP3P) potential, and a combined QM/MM simulation was performed for another 25 ps to further equilibrate the system at 298 K. Overall, about 250 ps of SBMD simulation were executed in the equilibration stage for the Michaelis complex.

In all calculations, the molecular dynamics trajectories were propagated with the leapfrog integrating scheme with a time step of 1 fs. All bond lengths and angles associated with hydrogen atoms were constrained by the SHAKE procedure.[60] A spherical electrostatic cutoff of 12.5 Å coupled with a switching function was used in this study, and the nonbonded pair list was updated in every 25 integration steps. The coordinates were saved at intervals of 100 steps during the simulation for additional analysis of the simulation trajectory.

2.5. Adaptive Umbrella Sampling. A principal goal of the present study is to distinguish concerted and stepwise mechanisms for each cyclization step by analyzing multidimensional potentials of mean force (PMF), $\Delta W(X_1, X_2)$, as a function of the relevant reaction coordinates, X_1 and X_2. The reaction coordinates for each ring formation will be specifically defined below. The potential of mean force was determined by an adaptive umbrella sampling technique that was implemented into CHARMM,[61] and it is related to the probability distribution $\bar{\rho}^o(X_1, X_2)$ averaged over all simulations:[62]

$$\Delta W(X_1, X_2) = -RT \ln \bar{\rho}^o(X_1, X_2) \quad (1)$$

where $\bar{\rho}^o(X_1, X_2)$ is the probability of finding the system at (X_1, X_2). In practice, a series of molecular dynamics simulations are carried out with the inclusion of a biasing potential to allow adequate sampling of regions with high potential energies. The adaptive umbrella sampling method makes use of a biasing potential $V_{bias}(X_1, X_2) = -\Delta W(X_1, X_2)$ that mirrors the computed potential of mean force and is continuously updated during the simulation.[61,63,64] Thus, as the simulation length increases, the sampling becomes more evenly distributed due to the use of improved accuracy in the biasing potential, and the computed PMF, $\Delta W(X_1, X_2)$, becomes better converged. Since we are not interested in regions of the potential energy surface away from the valleys along the reaction path, we employed an additional harmonic potential to restrain the range of distribution of the two reaction coordinates (X_1^i, X_2^i).[61] The effects of the biasing potential and the harmonic restraints are removed by the weighted histogram analysis method (WHAM) to obtain the unbiased distribution function, $\bar{\rho}^o(X_1, X_2)$.[65,66]

The two-dimensional free energy surfaces were obtained by the following procedure. We first carried out a set of 20 molecular dynamics simulations only using the harmonic restraining forces, centered equally along a diagonal line from the reactant (3.0, 3.0) to the product (1.5, 1.5) coordinates for the A/B ring formation step. As it turns out, none of the reaction steps followed strictly along this straight-line path; in fact, some processes are clearly stepwise reactions. This was readily identified during the simulation, and the centers of the harmonic restraints were adjusted to accommodate the curvature of the reaction path. Thus, the distribution in reaction coordinate sampling quickly settled into the valley regions of the low free-energy path for each system. The biased distribution of the reaction coordinates was obtained numerically by using bins of an area of 0.1×0.1 $Å^2$. Each simulation included 5 ps of equilibration and 10 ps of data collection, which yielded an initial guess of the potential of mean force and the biasing potential with WHAM.[65,66]

Next, three sets of 20 simulations, each of which includes 5 ps of equilibration and 20 ps of data collection, were run sequentially by including the biasing potentials and the harmonic restraints with their centers of the harmonic restraints placed along the minimum free-energy path. Of course, both the biasing potential and the free energy path were continuously being updated after each calculation in the adaptive umbrella sampling scheme. The simulations for each free energy surface were stopped until the computed PMF had changes less than 1% after an additional 100 ps molecular dynamics sampling. In all, each free energy surface was obtained from a total of 1.5 ns of simulation (time step of 1 fs).

The final two-dimensional PMFs were computed by regrouping all the data in bins of an area of 0.02×0.02 $Å^2$ to increase the accuracy in the computed reaction coordinate. However, smaller bin sizes lead to larger fluctuations and uncertainties in the average density distribution and computed free energy. In the present study, we found that there is no change in the computed barrier height and location of the transition state when the data were analyzed with bin sizes of 0.04×0.04 $Å^2$. Thus, we estimate that the statistical uncertainties in geometrical reaction coordinate are about 0.01 Å.

3. Results and Discussion

3.1. General Consideration. The strategy that we have undertaken is divide-and-conquer. Since there are five new C−C bonds formed in the squalene cyclization reaction in SHC, which is further complicated by the possibility of two five- to-six-membered ring expansions, it is not possible to describe the entire free energy surface in a simple one- or two-dimensional representation, nor is it feasible to determine a full five-dimensional free energy hypersurface. Thus, we have divided the reaction sequence into three major components, on the basis of information from mutagenesis experiments and analysis of products that have been trapped in these studies. We first examine the formation of A- and B-rings, with emphasis on the concertedness of A/B-ring formation (Scheme 2). Both A- and B-ring intermediates are tertiary cyclohexyl cations, whose formation can be monitored by the C_2–C_7 distance for A-ring and the C_6–C_{11} distance for B-ring, and a two-dimensional free energy contour as a function of $R^A_{C_2-C_7}$ and $R^B_{C_6-C_{11}}$ can afford an answer to this question.

Throughout this paper, we use a superscript to specify the ring with which the coordinate is associated and a subscript to indicate the definition of the coordinate, in terms of either the distance between two atoms or the difference in distance between two pairs of atoms. Therefore, the reaction coordinate $R^A_{C_2-C_7}$ denotes that the distance between C_2 and C_7 carbon atoms of squalene is used to describe the formation of A-ring.

(60) Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−337.
(61) Rajamani, R.; Naidoo, K.; Gao, J. *J. Comput. Chem.* **2003**, *24*, 1775−1781.
(62) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum: New York, 1977; Vol. 5, pp 169−194.
(63) Friedman, R. A.; Mezei, M. *J. Chem. Phys.* **1995**, *102*, 419−426.
(64) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865−880.
(65) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(66) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.

The free energy corresponding to the cation, **I**, has been set as the reference state, having a relative free energy of 0 kcal/mol in all figures.

The second step in our analysis is concerned with C- and D-ring formation, which formally undergoes two anti-Markovnikov additions (Scheme 2). However, the corresponding cyclopentylcarbinyl cations are thermodynamically favored, and the observation of aborted cyclization products that contain 6,6,5-fused and 6,6,6,5-fused ring skeletons has been used as strong evidence of the presence of these cationic intermediates.[6,7] It can be envisioned that the reaction proceeds first by forming the five-membered C-ring carbinyl cation, **III**, followed by an alkyl group migration from C_{14} to C_{15} assisted by the addition of the resulting secondary carbocation to the 18,19-olefin to produce the D-ring cyclopentylcarbinyl cation, **VI** (Scheme 2).[7] A relevant question is whether C-ring expansion and D-ring cyclization are concerted. For the five- to six-membered C-ring expansion, the distance between C_{10} and C_{15} of squalene gives a direct measure of the six-membered C-ring formation, and thus $R^{C}_{C_{10}-C_{15}}$ is chosen as the reaction coordinate for the C-ring expansion (formation) to form the cyclohexanyl cation **V**. The anchimeric assisted C-ring expansion involves major changes in two bond distances, the C_{10}–C_{14} bond distance and the C_{14}–C_{18} distance, which is associated with the formation of the D-ring cyclopentylcarbinyl cation **VI**. The difference between these two distances, i.e., $R^{C/D}_{C_{10}-C_{14},C_{14}-C_{18}}$, is used to specify the C-ring expansion and the anchimeric assisted addition to the 18,19-olefin.

The final ring closure gives the hopanyl cation intermediate **VIII**, which is the end-product of the cyclization reaction in SHC (Scheme 2). Unlike the OSC reaction, which involves a series of 1,2-shifts after the cyclization, the hopanyl cation is deprotonated by a basic residue, yet to be identified, or by a water molecule to afford the major product hopene in about 80% yield. In SHC, hopnayl cation **VIII** can also undergo nucleophilic addition by a water molecule to produce diplopterol in about 20% yield. The final E-ring cyclization requires D-ring expansion from the cyclopentylcarbinyl cation species, **VI**, if it is produced as a discrete intermediate along the squalene-to-hopene cyclization pathway. Alternatively, D- and E-ring formation may both be concerted along with the C-ring expansion process. Therefore, we choose the coordinate $R^{D}_{C_{14}-C_{19}}$ to monitor the progress of six-membered D-ring formation, similar to the C-ring formation step, and the coordinate $R^{D/E}_{C_{14}-C_{18},C_{18}-C_{22}}$ to specify the possibility of anchimeric D-ring expansion from the cyclopentylcarbinyl intermediate **VI** accompanied by the addition to the 22,23-olefin to form E-ring. In all, we construct three 2-dimensional free energy profiles to assess various reaction pathways mentioned above.

3.2. A- and B-Ring Formation. The cyclization reaction of squalene in SHC is initiated by protonation at the 2,3-olefin by the acidic residue Asp376, producing the putative tertiary C_2 carbocation, which undergoes addition to the 6,7-olefin to produce a tertiary carboncation intermediate **II**. In the reaction catalyzed by OSC in which 2,3-oxidosqualene is the substrate, Corey and Stass[22] established that the oxirane ring opening and A-ring formation are concerted, which is further supported by computational studies using ab initio molecular orbital and density functional theory.[12] Corey and Stass[22] showed that the rates for the proton-catalyzed cyclization reactions of **12** and

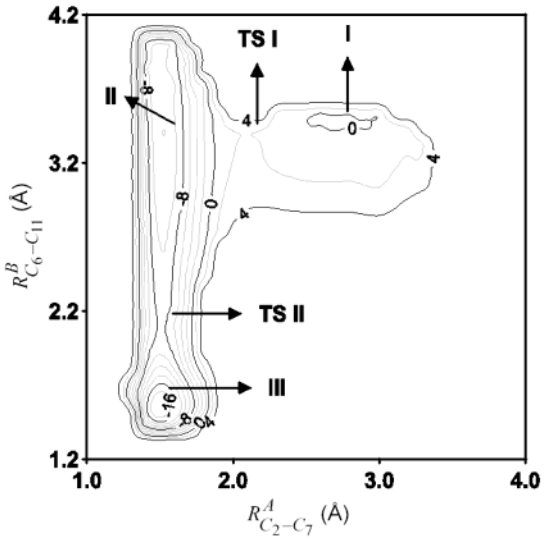
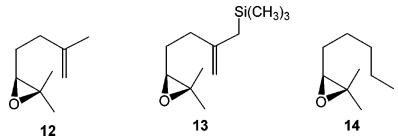


Figure 4. Computed two-dimensional free energy surface for the A- and B-ring formation. Geometrical parameters are given in angstroms and free energies are given in kilocalories per mole. The free energy for the carbenium ion **I** is set to zero and the contour levels are 2 kcal/mol.

Chart 1. Experimental Evidence for the Concerted Oxirane Opening and A-Ring Closure by Oxidosqualene−Lanosterol Cyclase[22]



13 were significantly greater than that of the carbocation formation from the noncyclizable species **14**, and the C−O cleavage was more favorable due to anchimeric assistance from the more nucleophilic olefin **13** than from **12** (Chart 1). It was suggested that the preorganized prechair conformation of the A-ring, coupled with a well-positioned acidic group (Asp456 in OSC), could produce rapid catalytic cation initialization. Since the oxirane species is much more basic than an olefin, anchimeric assistance from the 6,7-olefin is almost certainly essential to initiate the cyclization cascade in SHC.[6] The squalene protonation process is the rate-limiting step in the SHC reaction, although the mechanism of this reaction is deferred to a later publication. Here, we assume that the squalene substrate is already protonated by Asp376, and our investigation begins with the putative C_2 carbenium ion as the starting intermediate.

The free energy contour for the formation of rings A and B is shown in Figure 4 as a function of the two carbon−carbon bond distances, $R^{A}_{C_2-C_7}$ and $R^{B}_{C_6-C_{11}}$, that give rise to A- and B-ring closure. Cation **I** undergoes ring closure to form the A-ring cyclohexyl cation intermediate **II**, with a marginal barrier of 2.3 kcal/mol, which is consistent with the proposal that protonation and A-ring formation are concerted.[6] The intermediate **II** is −11.8 kcal/mol lower in free energy than the acyclic cation **I**. The conversion from **II** to the bicyclic cyclohexyl cation intermediate **III** has a computed free energy barrier of 3.9 kcal/mol, far smaller than the free energy gained in the formation of ring A. Even if energy dissipation is rapid in the active site of SHC, the lifetime of the A-ring intermediate **II** is expected to be short in the wild-type enzyme.

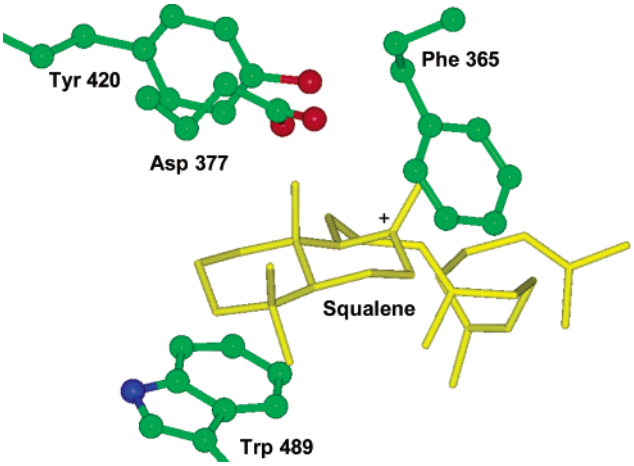
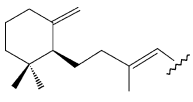


Figure 5. Snapshot of the structure obtained from molecular dynamic simulations corresponding to the A/B-bicyclic carbenium ion intermediate **III**.

Figure 4 reveals that the formation of rings A and B in the squalene cyclization process is a stepwise process, with the formation of a discrete A-ring carbocation intermediate. Nevertheless, only a small free-energy barrier of about 4 kcal/mol separates **II** and **III**. The net exogonicity from **I** to **III** is −19.3 kcal/mol due to the formation of two carbon−carbon bonds. The energy released in the two cyclization steps is consistent with previous theoretical studies of a model carbocation addition to form a tertiary cation product (the addition of 2-methyl-2-propyl cation to 2-methylpropene)[11] and the model cyclization reaction presented in Table 1. For the acyclic model system, the reaction proceeds without activation barrier with an enthalpy of reaction of −17.6 kcal/mol at the B3LYP/6-31G(d) level of theory, which is in accord with the gas-phase experimental value of −16.2 ± 1.6 kcal/mol. In methanol and THF solutions, a small solvent-induced barrier of about 3−4 kcal/mol was found from Monte Carlo simulations by Jenson and Jorgensen,[11] and there is only a small solvent effect on the reaction exothermicity. The model cyclization reaction corresponding to the formation of A-ring is also barrierless, and the predicted enthalpy of reaction in the gas phase is −13.7 kcal/mol with B3LYP/6-31G(d). The enthalpy difference between the acyclic and cyclic electrophilic addition was attributed to ring-strain effect.[11] Notice that B-ring formation is substantially less exogonic in Figure 4, which may be attributed to the extra ring strain effects to in the fused bicyclic system.

It has been proposed that Asp377, which is the third acidic residue in the DXDDTA motif, responsible for the initiation of the cyclization cascade in the catalytic site of squalene cyclase, provides electrostatic stabilization to the carbenium ion intermediates during the formation of A- and B-rings (Figure 5).[7,17] Mutations of D377N or D377C resulted in the formation of the aborted cyclization product **15** in 90% yield.[18,67] There is, however, concern about this proposal because a covalent enzyme−substrate complex may be formed from cation−anion recombination.



15

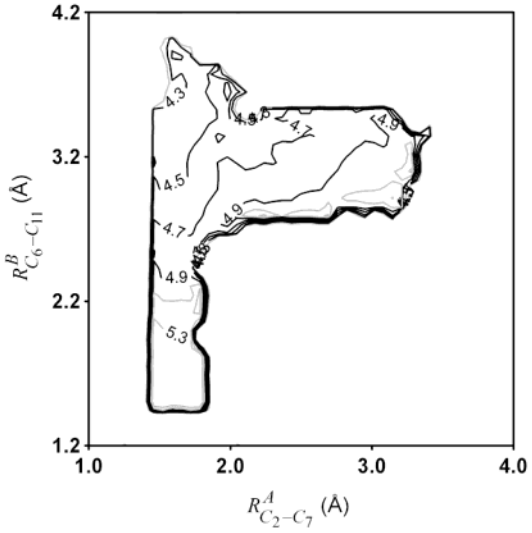


Figure 6. Average interatomic distances (angstroms) between the C_6 carbon of the substrate and the $O^{\delta 1}$ atom of Asp377 as a function of the two reaction coordinates defined in Figure 4. Contour levels are shown at 0.1 angstrom intervals.

To provide insight into the role of Asp377 on C_6 carbocation stabilization in the initial A-ring formation, we have computed the average distance between the oxygen atom ($O^{\delta 1}$) of Asp377 side chain and the substrate C_6 atom during the molecular dynamics simulation, and the interaction distance is plotted in Figure 6 as a function of the two reaction coordinates (see Figure 4). In regions surrounding cations **I** and **III**, the C_6–$O^{\delta 1}$ distance is about 5 Å or more, whereas in the area corresponding to the cyclohexyl cation **II**, where the A-ring cation is located on C_6, the C_6–$O^{\delta 1}$ distance shortens to about 4.1−4.3 Å due to increased electrostatic attractions. Thus, the development of the cationic charge at the C_6 position does indeed enhance electrostatic interactions with Asp377. To address the question of possible cation−anion combination in the active site and whether the computed average distance of 4.3 Å is an artifact of the QM/MM model, we have examined the electrostatic interaction of an acetate ion, treated by the MM-CHARMM force field, and *tert*-butyl cation, treated by the QM-AM1 model. The use of a combined QM/MM potential does not allow the formation of a covalent bond,[38,40] and the minimum distance between the two ions is dictated by the van der Waals parameters. We obtained an interaction energy of −98 kcal/mol and an interaction distance of 2.96 Å for this model system, far less than the average minimum distance of 4.3 Å. This indicates that the folding of the substrate and steric effects in the active site prevent Asp377 from getting too close to the C_6 cation but allow a shortening of the interaction distance by about 0.7 Å to provide stabilization as the cation is developed. This may also contribute to the lowering of barrier height in the cation initialization step.

The results from site-directed mutations suggest that steric interactions between the substrate and aromatic and aliphatic side chains play a crucial role in controlling the folding of the substrate and its subsequent cyclization cascade.[68] For example, the substitution of Tyr420 by a bulkier Trp residue leads to an altered preboat initial fold that gives rise to an aborted monocyclic product, whereas the less congested Tyr420Ala and

(67) Dang, T.; Prestwich, G. D. *Chem. Biol.* **2000**, *7*, 643−649.
(68) Full, C.; Poralla, K. *FEMS Microbiol. Lett.* **2000**, *183*, 221−224.

Tyr420Gly mutations yield bicyclic and tricyclic products from squalene (Figure 5).[23,69,70] Figure S1 in the Supporting Information displays the interaction distance between the phenolic hydroxyl oxygen and C_6 of the substrate during A/B-ring formation. The distance is essentially unchanged during the first ring closure, and then it increases by 0.9 Å to about 4.6 Å in the B-ring intermediate **III**. This indicates that the function of Tyr420 is not to stabilize the carbocation intermediates because the interaction distance would have otherwise shortened in A-ring formation, but it is consistent with the hypothesis of steric control. The smaller residues in the Tyr420Ala and Tyr420Gly mutations, which contain no cation$-\pi$ interactions with the substrate, created a larger active-site cavity that enables the stabilization of the B-ring (C_{10} cation) intermediate, perhaps by water molecules or the backbone carbonyl groups, to afford the bicyclic and tricyclic products.[69,70] On the other hand, the bulkier Trp mutant clashes with B-ring formation since it cannot effectively move away from the substrate as B-ring is formed, resulting in the isolation and accumulation of the monocyclic product.[70]

Site-directed mutagenesis studies indicated that Phe365 also plays a crucial role in the formation of B-ring and its further propagation in the cyclization reaction (Figure 5).[71] The F365A mutation resulted in 96% yield of the aborted 6,6-bicyclic products.[71] It has been suggested that the loss of cation$-\pi$ interactions, which would destabilize the carbocation and facilitate the elimination of protons in β-positions, is responsible for the experimental observation.[6,7,69,71] This proposal requires the presence of a base or nucleophile for the proton abstraction and that the proton abstraction rate is greater than cyclization. However, it is not clear which residue may serve as a viable base since the active site is aligned with aromatic and nonpolar residues. Furthermore, a destabilized carbocation intermediate would also lower the barrier and increase the rate of the cyclization steps. Thus, a destabilizing effect on intermediate **III** in the mutants would not have necessarily favored the formation of the bicyclic products. On the contrary, stabilization of **III** would have extended its lifetime in the active site, which would also increase the barrier to the next cyclization step to allow sufficient time for the deprotonation reaction to take place. Thus, we propose that steric effects between the substrate and Phe365, as Tyr420, which enforce substrate folding and prevent nucleophilic attack, are more important than cation$-\pi$ stabilization in balancing the stability of **III** as a key reaction intermediate in the cyclization cascade. This, nevertheless, does not preclude the importance of interactions between an aromatic π-system and the carbocation intermediates. The present simulation results show that the average distance between the center of the aromatic side chain and C_{10} decreases from about 5 Å in regions corresponding to intermediates **I** and **II** to about 4.6 Å in the region for the **III** minimum, suggesting that there is also substantial cation$-\pi$ interaction with Phe365. This is consistent with the observation that k_{cat} in the Phe365Tyr mutant increased 41-fold over that of the wild-type enzyme without any aborted reaction byproduct.[71]

3.3. Thermodynamic Control and the Formation of C- and D-Ring Intermediates. As it will become clear in the following discussion, the A/B-bicyclohexyl cation **III** is a key intermediate that branches the outcome of the cyclization process. This is because the formation of C- and D-rings in the squalene cyclization involves formally two anti-Markovnikov additions. In both cases, the Markovnikov intermediates are the thermodynamically favored, tertiary cyclopentylcarbinyl cations. Thus, the A/B-bicyclic cation **III** can undergo direct electrophilic addition to the 14,15-olefin to produce the thermodynamically favored 6,6,5-fused cyclopentylcarbinyl cation intermediate **IV**. Then, the five-membered C-ring goes through a ring expansion step to form the cyclohexyl cation intermediate **V**, which in turn adds to the 18,19-olefin to give the 6,6,6,5-tetracyclic carbinyl intermediate **VI**. These two steps can be either concerted or stepwise, which represents an important mechanistic question in understanding the cyclization process and in explaining the observation of byproducts in various mutations. To make computations tractable, in this section, we examine the preference for the formation of the five-membered (Markovnikov) and six-membered (anti-Markovnikov) C-ring intermediates. For the latter, we also include the anchimeric assisted electrophilic addition to form the tetracyclic D-ring **VI** to address the possibility that the anti-Markovnikov C-ring intermediate (**V**) is not a stable species.

Figure 7 shows the free energy contour as a function of the two geometrical variables $R^{C}_{C_{10}-C_{15}}$ and $R^{C/D}_{C_{14}-C_{18},C_{10}-C_{14}}$. Before we proceed, we first describe the structural features represented in Figure 7. The minimum near a coordinate at (0.8, 2.7) in the upper central region of Figure 7 represents the A/B-bicyclic cation **III**, and its energy level is anchored to be the same as that for **III** in Figure 4, which is -19.3 kcal/mol relative to **I**. We note that the distance between C_{10} and C_{15} atoms, which is the ordinate of Figure 7, provides a direct measure of the formation of the six-membered C-ring, which corresponds to a bond distance of about 1.55 Å. The abscissa coordinate specifies the anchimeric stabilization of the C-ring carbocation by the 18,19-olefin. Thus, if a 6,6,6-tricyclic system is produced as a stable intermediate, it would be shown in Figure 7 as a minimum with $R^{C}_{C_{10}-C_{15}} = 1.55$ Å. The second coordinate $R^{C/D}_{C_{14}-C_{18},C_{10}-C_{14}}$ will have either a positive or a negative value, depending, respectively, on whether the 18,19-olefin simply provides anchimeric stabilization to the secondary cation **V** or participates in electrophilic addition to yield the five-membered D-ring. The formation of the five-membered C-ring, cyclopentylcarbinyl cation **IV**, corresponds to a shortening of the C_{10}–C_{14} distance, with a nearly constant value of $R^{C}_{C_{10}-C_{15}} = 2.5$ Å, which is located at the upper right corner in Figure 7.

The most significant result of Figure 7 is that the secondary, carbocation species, corresponding to the six-membered C-ring *without* simultaneous formation of D-ring, was found to be much higher in energy than the Markovnikov product (**IV**) in the C-ring cyclization step. In fact, it cannot be located as a minimum on the free energy surface with positive values in $R^{C/D}_{C_{14}-C_{18},C_{10}-C_{14}}$ (large C_{14}–C_{18} distance) (Figure 7). This demonstrates that the anti-Markovnikov cyclization product is not a viable intermediate in the squalene cyclization cascade in the active site of SHC. This conclusion is consistent with the experimental observation that the 6,6,6-fused tricyclic product

(69) Pale-Grosdemange, C.; Merkofer, T.; Rohmer, M.; Poralla, K. *Tetrahedron Lett.* **1999**, *40*, 6009−6012.
(70) Sato, T.; Sasahara, S.; Yamakami, T.; Hoshino, T. *Biosci., Biotechnol., Biochem.* **2002**, *66*, 1660−1670.
(71) Hoshino, T.; Sato, T. *Chem. Commun. (Cambridge, U.K.)* **1999**, 2005−2006.

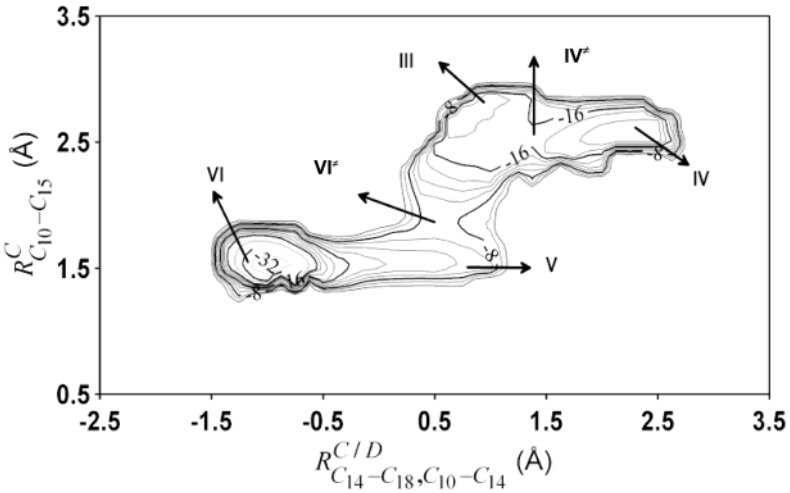


Figure 7. Computed free energy surface associated with the formation of C- and D-rings for the squalene cyclization in squalene−hopene cyclase. Free energies are relative to the carbenium ion **I**, and contour levels are shown at 2 kcal/mol intervals, except in the region near **VI**, where the contour levels are shown at 4 kcal/mol intervals for clarity.

has never been isolated, despite extensive efforts in mutagenesis studies.[6,7]

For the formation of the thermodynamically favored cyclopentylcarbinyl cation intermediate **IV** from **III**, the calculated free energy of activation is 2.9 kcal/mol, with a predicted free energy of formation of −1.8 kcal/mol relative to **III**. Previously, Jenson and Jorgensen[11] found that the formation of a cyclopentylcarbinyl cation from the corresponding acyclic cation releases −8.1 (−3.6) kcal/mol in enthalpy (free energy). The smaller value in the present study is due to the additional strain in the A/B/C-tricyclic ring skeleton compared to the monocyclic model compound. The reverse reaction from **IV** to **III** is also predicted to be rapid, with a barrier of only 4.7 kcal/mol. Thus, **III** and **IV** should be in rapid thermodynamic equilibrium, and if mutations are made to stabilize the cyclopentylcarbinyl cation **IV**, it would have a sufficient lifetime to be trapped at this stage to produce the 6,6,5-tricyclic products resulting from elimination of a proton. Indeed, numerous 6,6,5-tricyclic byproducts have been isolated, particularly in mutations of amino acid residues in close proximity to the C-ring region,[7] including the F601A and F605A replacements, which are located near ring C of the hopene skeleton (Figure 8).[72]

Since the existence of a discrete 6-membered C-ring intermediate resulting from ring expansion can be ruled out in view of the free energy profile in Figure 7, the only possibility for C-ring formation is a concerted C/D-ring cyclization. In this case, the D-ring intermediate can be either a thermodynamically favored cyclopentylcarbinyl cation, **VI**, or the six-membered anti-Markovnikov product, **VII**, and the distinction of these two reaction paths will be discussed in the next section. Figure 7 shows that an asynchronous concerted path indeed exists from the A/B-bicyclic intermediate **III** to the 6,6,6,5-tetracyclic fused carbocation intermediate **VI** with a free energy barrier of 10.9 kcal/mol. The barrier can be attributed to the energy cost for the formation of the six-membered C-ring, which is readily compensated by the stabilization in the formation of the tertiary D-ring carbenium ion. Remarkably, **VI** is predicted to be −18.7 kcal/mol exogonic relative to **III**, significantly more stable than

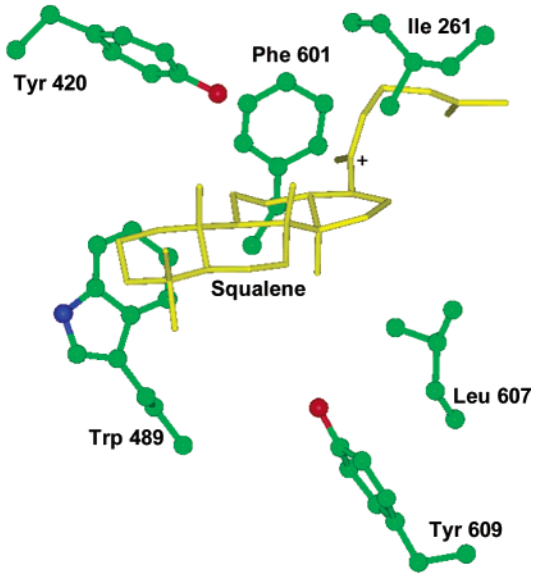


Figure 8. Snapshot of key residues near the 6,6,6,5-tetracyclic carbenium ion intermediate **VI** from molecular dynamics simulations.

the tricyclic intermediate **IV**. Figure 7 also shows that although the tricyclic intermediate **IV** is kinetically favored with a very low activation barrier of 2.9 kcal/mol, the thermodynamic competition leading to the formation of the tetracyclic species, **VI**, can be strongly preferred, especially in view of the fact that **III** and **IV** would be in thermodynamic equilibrium under conditions leading to the formation of **VI**.

Experimentally, numerous aborted cyclization products with the 6,6,6,5-fused tetracyclic skeleton have been isolated, both in the wild-type enzyme (with about 1% of total yield) and in numerous mutants incubated with squalene. This, along with the finding of a significant energy preference in Figure 7, seems to strongly indicate that **VI** is a key intermediate in the squalene-to-hopene cyclization cascade in SHC. Indeed, the five- to six-membered D-ring expansion process has been a key feature in all proposed mechanisms.[5−7,25] However, we show in the following section that this does not appear to be feasible because the barrier for the D-ring expansion is as large as the barrier in the **VI**-to-**III** back reaction, which is 29.6 kcal/mol. It might

(72) Hoshino, T.; Kouda, M.; Abe, T.; Sato, T. *Chem. Commun. (Cambridge, U.K.)* **2000**, 1485−1486.

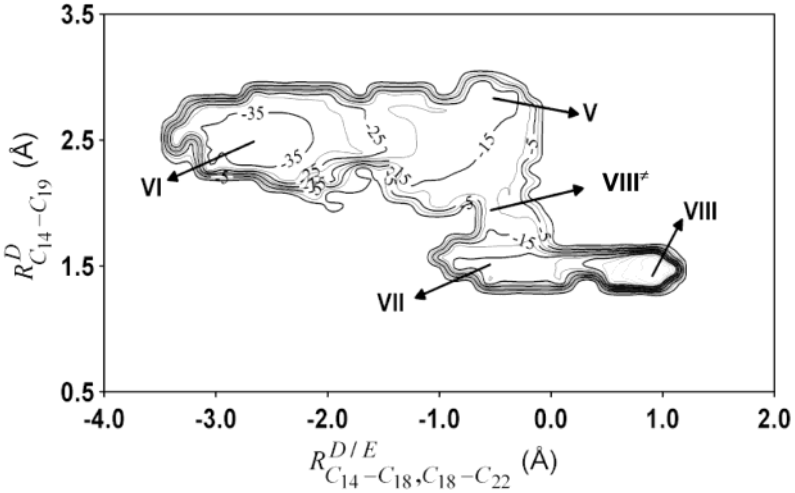


Figure 9. Computed free energy surface for the five- to six-membered D-ring expansion and E-ring formation in squalene−hopene cyclase. Free energies are relative to the carbenium ion **I**, and contour levels are shown at 5 kcal/mol intervals.

be tempting to suggest that the energy released from the ring formation steps can be used to overcome the kinetic barrier, but we expect that energy dissipation is rapid and that it will not be stored for an extended period of time in the enzyme comparable to the chemical process ($k_{cat} = \sim 6$ min^{-1} at 30 °C).[7] Consequently, the 6,6,6,5-fused tetracyclic cation intermediate **VI,** in fact, would represent a thermodynamic trap in the squalene cyclization pathway. Once **VI** is formed, it will not undergo ring expansion to produce the hopanyl cation, but it will lead to the formation of byproducts of the wild-type enzyme. Therefore, the enzyme squalene cyclase must use another control to avoid the formation of the 6,6,6,5-cyclopentylcarbinyl cation intermediate **VI**, but 1% of the cyclization process still fall into this thermodynamic well.

3.4. Kinetic Control and the Formation of Hopanyl Cation. The 6,6,6,5-tetracylic cyclopentylcarbinyl cation **VI** modeled in Figure 7, if formed as an "on-pathway intermediate",[6] must expand to yield the six-membered D-ring, which undergoes the final cyclization step to form the final hopanyl cation **VIII** (Scheme 2). Similar to the C-ring expansion process, the D-ring expansion can be characterized by a single coordinate, the distance between C_{14} and C_{19} carbon atoms, $R^{D}_{C_{14}-C_{19}}$. This corresponds to a change in $R^{D}_{C_{14}-C_{19}}$ from 2.5 to about 1.5 Å. Anchimeric assistance to the D-ring expansion along with E-ring formation is represented by the difference of two bond distances specified by the coordinate $R^{D/E}_{C_{14}-C_{18},C_{18}-C_{22}}$. The free energy contours for these processes are shown in Figure 9, and contour levels have been anchored at the minimum of cation **VI** with respect to the value (−38.0 kcal/mol) in Figure 7.

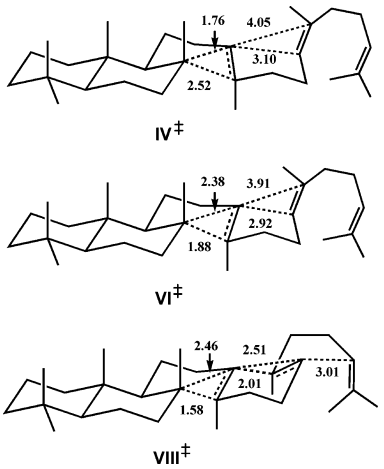
First, Figure 9 shows that a stepwise D-ring expansion mechanism is not energetically viable. This would correspond to a free energy minimum near (−2.5, 1.6) in Figure 9; however, this region was found to be very high in free energy, making it have little chemical consequence. An anchimeric assisted D-ring expansion pathway, which is highly asynchronous but concerted, was located in Figure 9 with a single transition state, **VIII**‡, at (−0.5, 1.9). The process can be described as an initial nucleophilic attack by the 22,23-olefin at the C_{19} cation, indicated by the change of the $R^{D/E}_{C_{14}-C_{18},C_{18}-C_{22}}$ coordinate, followed by alkyl group migration in the D-ring expansion, which is the vertical change in the $R^{D}_{C_{14}-C_{19}}$ coordinate in Figure 9. The process is completed by the final C_{18}–C_{22} bond formation in E-ring closure. The activation free energy for this process is estimated to be 27.8 kcal/mol, which is 1.8 kcal/mol lower than the back-reaction from **VI** to **III**.

This large free energy barrier represents a major problem in the squalene cyclization cascade because unless energy dissipation is extremely slow, such that the free energy gained in the formation of **VI** can be used to overcome the barrier of **VIII**‡, the rate-limiting step in the SHC reaction would be the D-ring expansion. Experimentally, the turnover number is 6.4 min^{-1} at 30 °C by the wild-type enzyme,[7] which is a much longer time scale than the energy dissipation rate in proteins.[73] The existence of a thermodynamic sink along an enzyme reaction pathway is also contradictory to the classic picture of multiple energy barriers and wells of similar magnitude that was proposed for perfection of enzymes by Knowles and Albery.[74] Thus, it appears that an alternative path must exist in the squalene cyclization cascade to avoid the thermodynamic trap of the carbocationic intermediate **VI**.

Some clues can be obtained by combining the free energy contours of Figures 7 and 9, and we summarize the computed average bond distances for key structural parameters in the three transition states, leading to carbocation intermediates **IV**, **VI**, and **VIII**, in Chart 2. Transition structure **IV**‡ is associated with the formation of the tertiary cyclopentylcarbinyl cation **IV**, which requires no assistance by anchimeric stabilization from the 18,19-olefin, and adopts one of the two asymmetric forms of the bridged carbocation (with a shorter C_{10}–C_{14} bond length at 1.76 Å). On the other hand, the formation of the six-membered C-ring must be stabilized by anchimeric assistance, giving rise to the other asymmetric form of the bridged cation transition structure (a shorter C_{10}–C_{15} distance at 1.88 Å) in **VI**‡. Furthermore, the anchimeric interaction is not symmetric, such that the C_{14}–C_{18} and C_{14}–C_{19} distances are 2.92 and 3.91 Å, respectively, and this leads to formation of the five-membered, tertiary carbocation **VI**. The anchimeric stabilization can also have a shorter C_{14}–C_{19} distance than the C_{14}–C_{18} separation due to a further stabilization of the putative carbocation at the C_{18} position by the 22,23-olefin π-bond. This

(73) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1−259.

Chart 2. Schematic Representation of the Three Transition-State Structures Leading to the Carbocation Intermediates **IV**, **VI**, and **VIII**, Respectively[a]



[a] Average distances for key geometrical variables are given in angstroms.

corresponds to the transition state (**VIII**‡) for the D-ring expansion process (Chart 2). Comparison of the two transition structures, **VI**‡ and **VIII**‡, shows that the two transition states are actually similar, but the latter consists of a more extensively folded conformation of the substrate, in which the C_{10}–C_{15} bond is essentially formed and the formally secondary cation at C_{18} is further stabilized by the 22,23-olefin.

The presence of an additional anchimeric stabilization by the 22,23-olefin, which also leads to E-ring formation, makes transition state **VIII**‡ 1.8 kcal/mol lower in free energy than **VI**‡. In view of the similarity between the two transition structures (**VI**‡ and **VIII**‡), we propose that the formation of all C-, D-, and E-rings from the carbenium ion intermediate **III** is concerted and that there is kinetic competition in wild-type SHC through transition state **VI**‡ and a **VIII**‡-like transition state. The former, which has been characterized in Figure 7, produces the 6,6,6,5-tetracyclic intermediate, giving rise to byproducts in about 1% yields, whereas the latter leads to the 6,6,6,6,5-pentacyclic products hopene and diplopterol (hopanol). The key feature in this proposal, different from all other proposed mechanisms,[5−7,25] is that the five- to six-membered D-ring expansion processes are not feasible reaction pathways because of the extremely high thermodynamic instability of intermediates **V** and **VI** (Figures 7 and 9) Thus, the squalene cyclization cascade proceeds through either a concerted C/D-ring formation, leading to the tetracyclic byproducts, or a concerted, but highly asynchronous, C/D/E-ring formation pathway to yield the "correct" products, and the observed product formation is controlled by the kinetic barriers from the A/B-bicyclic cation intermediate **III**.

Although it would be desirable to characterize explicitly the nature of the transition state of the extended concerted reaction pathway, which requires the construction of a three-dimensional free energy hypersurface, the high computational cost makes it intractable at the present time. Because of this difficulty, we emphasize that our proposal of the extended concerted formation of rings C/D/E is an extrapolation of the computed structural data (Chart 2) and experimental information. The conformational

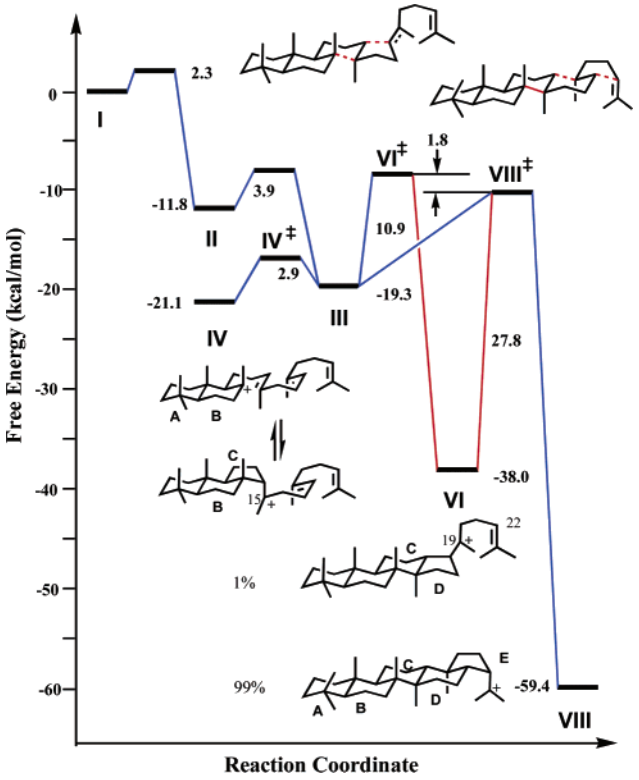


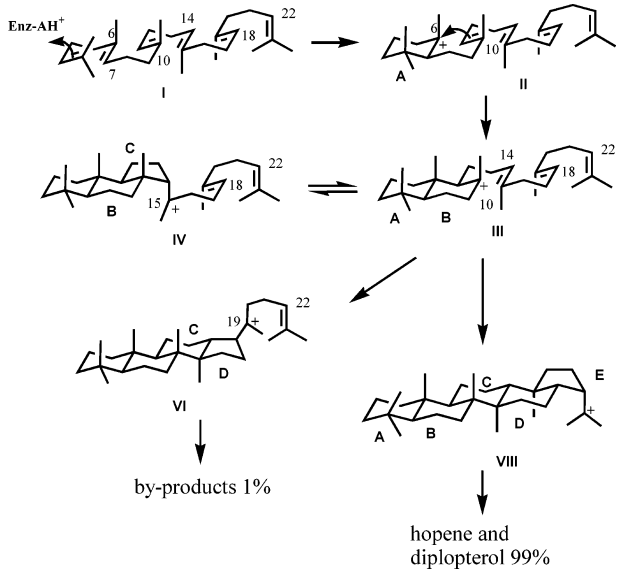
Figure 10. Relative free energies for key reaction intermediate and transition states in the mechanism of squalene cyclization (Scheme 5) in squalene−hopene cyclase.

fluctuation of the substrate in the active site, leading to the presence or absence of the 22,23-olefin anchimeric stabilization of the D-ring cationic intermediate, is a dominant factor to lower the energy of **VIII**‡. Clearly, the formation of the C_{10}–C_{15} bond is much more advanced in **VIII**‡ than in **VI**‡, and there could potentially exist a shallow minimum near this region of the potential energy surface. Nevertheless, **VIII**‡ is a good approximation of the transition state for this asynchronous transformation. If we make a further estimate using the free energy difference between **VI**‡ and **VIII**‡, we obtain a predicted ratio of 5/95 for branching into **VI** and **VIII**, which is in remarkable accord with experiment (a ratio of 1/99 in wild-type SHC).[25]

3.5. Mechanism. We summarize the energetic results from the two-dimensional free energy simulations in Figure 10, which shows the free energy barriers and the free energies of reaction for the formation of various stable carbenium ion intermediates in the squalene cyclization cascade by SHC. The reaction coordinate in Figure 10 can be considered as the minimum free energy path shown in Figures 4, 7, and 9, except in the connection from **III** to **VIII**‡, which is actually computed for the reaction from **VI** to **VIII**. This free energy diagram led us to propose a refined cyclization mechanism from squalene to hopanyl cation in wild-type SHC (Scheme 5). The protonation of squalene and A-ring formation are considered to be concerted, which is supported by the present calculation for the conversion of **I** to **II** with a negligible free energy barrier of 2.3 kcal/mol and a free energy of formation of −11.8 kcal/mol. Although the A-ring cyclohexyl cation **II** is a stable species, only a marginal free energy of activation of 3.9 kcal/mol separates it from the key bicyclic carbocation intermediate **III**. The second cyclization step gains −7.5 kcal/mol in free energy.

(74) Knowles, J. R.; Albery, W. J. *Acc. Chem. Res.* **1977**, *10*, 105−111.

Scheme 5. Proposed Carbocation Cyclization Mechanism in Squalene−Hopene Cyclase



Carbocation intermediate **III** can undergo three separate concerted transformations, to produce the tricyclic, tetracyclic, and pentacyclic cyclopentylcarbinyl cation intermediates **IV**, **VI**, and **VIII**, respectively, involving the formation of one, two, and three new rings and carbon−carbon bonds. Computational results in Figure 7 show that **III** and **IV** are nearly thermoneutral with **IV** being only 1.8 kcal/mol lower in free energy than **III**, and the barrier separating the two species is too small to have any significant kinetic impact on their interconversion. Thus, we propose that **III** and **IV** are in thermodynamic equilibrium under the conditions of carbocation cyclization in the active site of SHC. Significantly, there is no alternative path for **IV** to undergo further chemical transformation because free energy simulations revealed that it is not feasible energetically for **IV** to undergo a five- to six-membered ring expansion to yield the anti-Markovnikov C-ring cyclohexyl cation (**V**). Further, **IV** is not particularly more stabilized than **III** and the active site along the cyclization pathway consists of only aromatic and aliphatic residues, making elimination or addition reactions not easily accessible unless cation **IV** is *stabilized* as in site-directed mutations.[23,25,28,68−72,75−77] The formations of both the experimentally observed byproducts, corresponding to the 6,6,6,5-tetracyclic intermediate **VI**,[25] and the native products from the hopanyl cation skeleton (**VIII**) are thermodynamically favored and are considered to be highly asynchronous concerted processes. Note that the concertedness for formation of **VIII** is not directly validated in the present study (see discussion above). The branching ratio between **VI** and **VIII** is dictated kinetically by the difference between the two reaction barriers, and the present simulation study yields a value of 1.8 kcal/mol with **VIII**‡ lying lower in free energy than **VI**‡. Because of the thermodynamic stability of **VI**, we propose that ring expansion in going from **VI** to **VIII**, either through a discrete six-membered D-ring intermediate or an anchimeric assistance with spontaneous cyclization of the E-ring, is not likely, and **VI**, in fact, represents a thermodynamic trap along the cation cyclization pathway.

It is interesting to compare the delicate balance of thermodynamic and kinetic control of the reaction pathways from the carbocation intermediate **III**. To avoid stabilization of the tricyclic, cyclopentylcarbinyl cation **IV**, the enzyme appears to have raised its free energy to a level similar to that of **III** (recall that each cyclization step releases about 8−12 kcal/mol free energy),[11] making it readily converted back into **III** as the latter is consumed. The forward cyclization steps from **III** are dictated by thermodynamic stability in the formation of both D- and E-rings; however, steric strain is no longer able to keep the energy of tetracyclic cationic intermediate **VI** similar to that of **III**. Instead, the enzyme makes use of kinetic control in favor of the formation of the desired product, hopanyl cation **VIII**.

The active site at the C_{23} cationic position of **VIII** is occupied by a number of water molecules in addition to Glu45. The solvent water molecules and Glu45 can facilitate the deprotonation or addition reactions to yield hopene and diplopterol. Interestingly, there does not appear to be any specific control in the formation of the alkene and alcohol products by the enzyme.

4. Conclusions

Molecular dynamics simulations have been carried out to investigate the carbocation cyclization mechanism from squalene to hopene/diplopterol in the active site of squalene−hopene cyclase. The present study, which was based on free energy simulations by constructing the free energy surfaces for the cyclization steps along the reaction pathway, has addressed the detailed reaction mechanism associated with each ring closure, the stability of all possible carbenium ion intermediates, and the free energies of activation accompanying their interconversions. The principal findings are as follows.

(1) The squalene to hopene cyclization reaction in SHC is remarkably exogonic with a predicted free energy of reaction of 59 kcal/mol from the C_3 carbenium ion **I** to hopanyl cation **VIII**. If the initial protonation step accompanying A-ring formation is considered, the overall reaction is still expected to be exothermic by about 40−50 kcal/mol. Such a large exothermicity in an enzyme reaction is unusual, although it is consistent with previous computational results on intramolecular ring closure, releasing 8−12 kcal/mol.[11] It has been proposed that the five to seven repeated QW motifs, rich in aromatic residues, play an important role in reinforcing the integrity of the protein structure.[13]

(2) The anticipated five- to six-membered ring expansion processes that have been proposed for C- and D-ring formation in all previous cyclization mechanisms are not viable reaction pathways. The six-membered C-ring species is not a minimum on the free energy surface and its formation is accompanied by an asynchronous concerted cyclization of ring D. This is consistent with the experimental results that the 6,6,6-tricyclic aborted cyclization product has never been isolated. The formation of the Markovnikov cyclopentylcarbinyl D-ring intermediate represents a thermodynamic trap along the squalene to hopene cyclization cascade because once it is formed, it requires an activation free energy of about 28 kcal/mol in the rearrangement of the tertiary cation **VI** to the final hopanyl cation **VIII**. Indeed, numerous tetracyclic byproducts, resulting

(75) Hoshino, T. *Nippon Nogei Kagaku Kaishi* **2002**, *76*, 1187−1190.
(76) Sato, T.; Hoshino, T. *Biosci., Biotechnol., Biochem.* **2001**, *65*, 2233−2242.
(77) Hoshino, T.; Ohashi, S. *Organic Lett.* **2002**, *4*, 2553−2556.

from the carbenium ion **VI** intermediate, have been isolated in wild-type and mutant enzymes.

(3) The bicyclic carbenium ion **III** is the only reaction intermediate with a significant lifetime along the squalene cyclization pathway. It appears that squalene cyclase avoids the Markovnikov addition intermediate **IV** from being trapped in the reaction pathway by raising its free energy to a level similar to that of **III** such that these two species are in thermodynamic equilibrium. Previous model studies showed that the closure of an extended structure to yield the cyclopentylcarbinyl cation is exothermic by 8−9 kcal/mol, and it was suggested that the formation energy from a more preorganized conformer in the enzyme would be several kilocalories per mole more exothermic.[11] The present study shows that **IV** is only 1.8 kcal/mol lower in free energy than **III**, suggesting that the enzyme destabilizes the formation of **IV**.

(4) The formation of the final hopanyl cation **VIII** is achieved by a highly asynchronous concerted step involving the formation of rings C, D, and E, and the preference for this reaction path over that to yield **VI** is kinetically controlled by a lower free energy barrier.

(5) On the basis of the computational results, we propose the following mechanism on hopene biosynthesis (Scheme 5). Squalene is preorganized in an all-chair conformation state in the active site and the cyclization reaction is initiated by protonation from Asp376 accompanied by A-ring closure to yield **II**. Although **II** is a true minimum on the free energy surface, its lifetime is expected to be short because there is only a small activation barrier (3.9 kcal/mol) in the next addition to produce the key intermediate **III** in the reaction pathway. The second ring-closure step releases −7.5 kcal/mol in free energy. **III** can undergo three separate concerted transformations to yield the tricyclic, tetracyclic, and pentacyclic cyclopentylcarbinyl cation intermediates **IV**, **VI**, and **VIII**, respectively. Formation of **IV** does not lead to further additional reactions (except in mutant enzymes), but it is in rapid thermodynamic equilibrium with **III**. Therefore, its formation does not pose problems in the squalene cyclization cascade because **IV** is readily converted back into **III** as it is consumed in the other two processes. Concerted formation of the anti-Markovnikov C-ring and Markovnikov D-ring, **VI**, is thermodynamically favored by about 29 kcal/mol relative to **III** with a free energy of activation of 10.9 kcal/mol (Figure 10), but the kinetic barrier for the conversion into **VIII** is smaller by about 2 kcal/mol, making the formation of the final carbenium ion intermediate kinetically preferred.

The delicate balance of thermodynamic and kinetic control in the squalene to hopene cation cyclization is remarkable. To avoid stabilization of the tricyclic, cyclopentylcarbinyl cation **IV**, which would be a favored Markovnikov product, the enzyme has raised its free energy so that rapid equilibrium can be established with **III**. We expect that the lifetime for **IV** is similar to that for **III**; consequently, elimination and addition reactions of **IV** do not occur at this stage and the enzyme does not need to avoid its formation. The further cyclization steps from **III** are highly favored thermodynamically, but the enzyme controls the formation of **VIII** by a faster reaction rate than that for the byproduct **VI**.

Site-directed mutagenesis experiments probe specific interactions as a result of the residue mutation, leading to stabilization or destabilization of various carbenium ion intermediates relative to the wild-type enzyme. The isolation of a particular aborted cyclization byproduct is not necessarily a direct indication of the presence of the corresponding cationic intermediate along the reaction pathway in the wild-type enzyme. On the contrary, it is a testimony that these intermediates are actually avoided to be populated with extended lifetime in wild-type SHC. The mutation experiments do, however, prove the significance of these residues in regio- and stereochemical control in the active site if different products than those in the wild-type enzyme are obtained. While the present computation study has provided insights into the energetics and pathways and the stability of key carbenium ion intermediates in the squalene to hopene cyclization reaction, clearly, many details remain to be resolved, especially on the interpretation of site-directed mutagenesis results of this remarkable enzyme.

Acknowledgment. This work has been supported by the National Institutes of Health.

Supporting Information Available: One figure depicting the distribution of interatomic distances discussed in the text (PDF). This information is available free of charge via the Internet at http://pubs.acs.org.

JA0371799