JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Physical Constraints on Sesquiterpene Diversity Arising from Cyclization of the Eudesm-5-yl Carbocation

B. Andes Hess, Jr.,[*,†] Lidia Smentek,[†] Joseph P. Noel,[‡,§] and Paul E. O'Maille[*,∥,⊥]

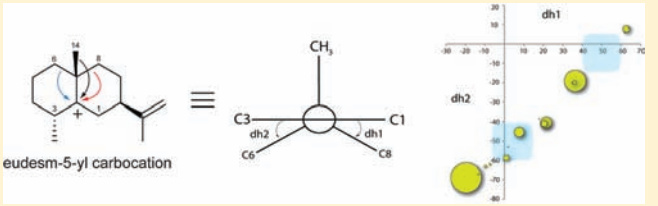
[†]Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States

[‡]Howard Hughes Medicinal Institute, and [§]The Salk Institute for Biological Studies, Jack H. Skirball Center for Chemical Biology & Proteomics, 10010 Torrey Pines Road, La Jolla, California 92037, United States

[∥]Department of Metabolic Biology, John Innes Centre (JIC), and [⊥]Phytochemicals and Health Programme, Institute of Food Research (IFR), Norwich Research Park, Norwich NR4 7UH, United Kingdom

Supporting Information

ABSTRACT: The biogenic origins of complex cyclic terpenes derive from the interplay of enzymes and the intrinsic reactivity of carbocation species at major branch-points along intramolecular cyclization pathways to ultimately determine the distribution of terpene skeletal types in nature. Solanaceous plants biosynthesize chemical defense compounds, largely derived from the eremophilane and spirovetivane-type sesquiterpenes. These hydrocarbon skeletons share a common biogenic origin, stemming from alternative Wagner−Meerwein rearrangements of the eudesm-5-yl carbocation during the cyclization of farnesyl pyrophosphate (FPP) catalyzed by sesquiterpene synthases. While the spirojatamane skeleton shares the same carbocation intermediate, this class of sesquiterpenes has not been reported in the *Solanaceae* and is exceedingly rare in nature. To investigate the physical basis for alternative rearrangements of the eudesm-5-yl carbocation, we carried out quantum mechanics (QM) analyses to calculate the allowable conformations, energies, and transition states linking conformers of the eudesm-5-yl carbocation to the eremophilene, spirovetivane, and spirojatamane skeletons. Additionally, we conducted parallel investigations on simplified decalin carbocation systems to examine the contribution of ring substituents to allowable conformations and rearrangement pathways. Our study reveals that ring substituents expand the conformational space accessible to the eudesm-5-yl carbocation while sterically blocking rearrangements in certain contexts. From our analysis, we define a conformational threshold for each possible rearrangement based on dihedral angles describing transition state geometry. Further, our calculations indicate that methylene migration rearrangements leading to spiro compounds are thermodynamically dominant in the eudesm-5-yl and simpler decalin carabocation systems. Interestingly, the theoretical abundance of sesquiterpene skeletal types arising from the intrinsic reactivity of the eudesm-5-yl carbocation stands in sharp contrast to their currently known natural abundance. The implications of these results for the catalytic tragectories catalyzed by sesquiterpene synthases are discussed.



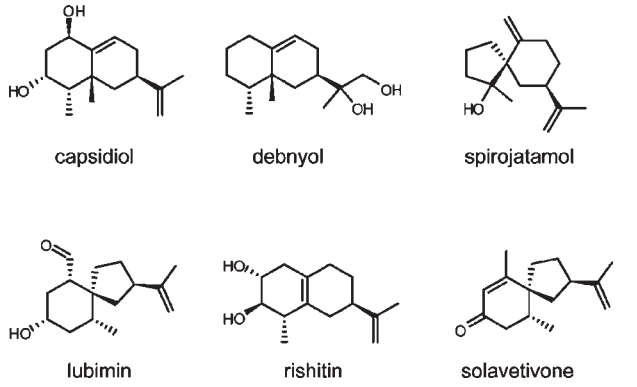
INTRODUCTION

Terpenes are the largest and most structurally complex group of plant natural products, numbering well over 30 000 known compounds.[1] These metabolites are essential for plant ecological viability[2−7] while providing important bioactives for human uses.[8,9] Remarkably, the stereochemical complexity of terpene hydrocarbon skeletons largely derives from a single enzyme-catalyzed reaction by terpene synthases. For example, sesquiterpene synthases may produce one of 300 stereochemically distinct hydrocarbon skeletons from a universal substrate farnesyl diphosphate (FPP) (**1**) (for review[10]). More generally, terpene synthases generate carbocations via metal-assisted and/or protonation mechanisms, and then direct a sequence of intramolecular cyclizations, often involving numerous Wagner−Meerwein rearrangements, terminating with charge neutralization by water capture and proton elimination or simply proton elimination. The intersection of an evolving terpene synthase and the intrinsic reactivity of its isoprenoid diphosphate substrate, particularly at critical carbocation branch-points along cyclization pathways, dictates final product outcomes and hence the distribution of terpene skeletal types in nature.[11] To what extent does terpene chemical diversity reflect the intrinsic reactivity of the carbocation? To address this fundamental question, this Article presents a detailed investigation of the conformation, energy, and reactivity of the eudesm-5-yl carbocation.

Carbocations in terpene cyclization reactions are transient intermediates normally recalcitrant to isolation. However, their existence has been well corroborated through careful isotopic labeling investigations[12−14] and theoretical studies from quantum chemistry (for review[15]). This forms the basis for widely accepted stereochemical rules of sesquiterpene biogenesis.[16]

Received: April 12, 2011
Published: June 29, 2011

Scheme 1. Representative Spirane and Eremophilane Sesquiterpene Natural Products



Accordingly, the attacking nucleophile must be perpendicular to the plane of the carbocation for carbon−carbon bond formation to occur. Therefore, knowing the stereochemistry of the products, one can apply this chemical logic to discern the likely structure and conformation of the carbocation intermediates leading to neutral end-product formation.[17,18]

In nature, the distribution and abundance of sesquiterpene skeletal types varies widely with some correlation to particular taxonomic groups. For example, chemical diversification of sesquiterpene lactones throughout *Asteraceae* tracks with known phylogenetic relationships, providing a set of chemosystematic markers routinely used to identify particular species.[19−21] Solanaceous plants produce a signature collection of sesquiterpene phytoalexins. For example, tobacco and peppers produce capsidiol[22] and debneyol,[23] derivatives of the eremophilane skeletal class, as their chemical defense compounds, while potato, tomato, and henbane produce lubimin,[24] rishitin,[25] and solavetivone,[26] among a suite of spirovetivane-derived compounds (Scheme 1). The spirojatamane skeletal type is a sister to both eremophilane and spirovetivane terpenes (discussed below); however, there are no reports of any derivatives in solanaceaous plants. In fact, spirojatamanes are a rare skeletal class first isolated from *Lepechinia bullata*[27] with their named based on the ring system of (+)-spirojatamol, a sesquiterpene alcohol of *Nardostachys jatamansi*.[28]

To investigate the evolutionary origins of eremophilane and spriovetivane terpene biogenesis, we constructed and analyzed the catalytic properties of a collection of phylogenetically defined mutants encompassing some possible evolutionary intermediates of *Nicotiana tabacum* (+)-5-*epi*-aristolochene synthase (TEAS) and *Hyoscyamus muticus* premanspirodiene synthase (HPS), an extant pair of closely related sesquiterpene cyclases from the *Solanaceae* with divergent catalytic properties. These studies revealed a rugged "catalytic landscape" where promiscuous catalytic behavior is prevalent in the majority of mutants, while specific activities are achieved by degenerate combinations of mutations, sometimes linked by single mutational changes.[29] On a mechanistic level, the observed catalytic properties across the TEAS- and HPS-like enzymes hinge upon the differential control of a common eudesm-5-yl carbocation intermediate (**2**), which undergoes endothermic Wagner−Meerwein rearrangements leading to predominantly 5-*epi*-aritolochene (**5EA**) or premnaspirodiene (**PSD**), to produce eremophilane and spirovetivane

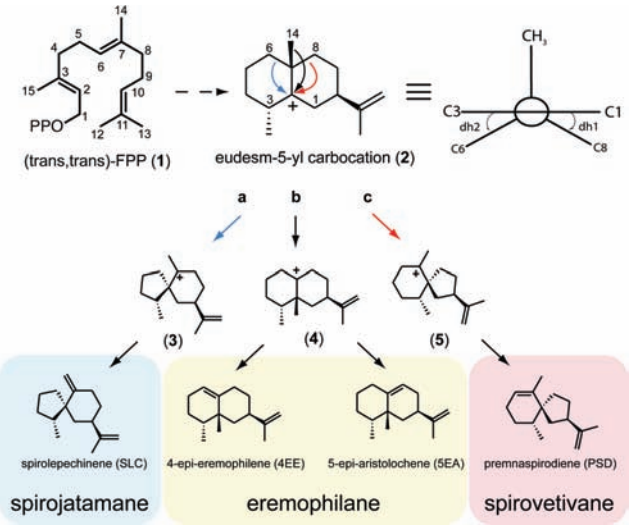


Figure 1. Mechanistic branch point leading to eremophilane, spirovetivane, and spirojamane cyclization products. Farnesyl pyrophosphate (FPP) is ionized and cyclized to the eudesm-5-yl carbocation (**2**) by terpene synthases from the *Solanaceae* plant family. Further Wagner−Meerwein rearrangement of this cation along path a, b, or c produces daughter cations **3**, **4**, and **5** leading to the spirojatamane, eremophilane, and spirovetivane sesquiterpenes skeletal classes, respectively. Viewed along the plane of the carbocation center (carbon 2), a Newman projection of the eudesm-5-yl carbocation (**2**) reveals a pair of dihedral angles formed by C1278 or C3276 defined as dh1 and dh2, respectively.

sesquiterpene skeletal types, respectively (Figure 1). In a few instances, 4-*epi*-eremophilene (**4EE**), an eremophilane, was observed as the dominant product, while the spirojatamane compound spirolepechinene (**SLC**), a previously identified minor product of TEAS from in vitro assays,[18] never emerged as the dominant product, despite its mechanistic descent from the shared parent eudesm-5-yl carbocation.

Observations from these in vitro experiments and from the currently known abundance of terpene skeletal types in plants raise intriguing evolutionary questions rooted in physical chemistry. Is there a geometric or energetic basis for the dominance of eremophilane and spirovetivane skeletal forms over that of spirojatamanes? What are the physical and chemical constraints of the eudesm-5-yl carbocation that dictate the conformation, energy, and accessible cyclization pathways? How closely does the theoretical abundance of cyclization products correlate with the natural abundance of these sesquiterpenes? Does natural selection follow the intrinsic physical-chemical constraints of the intermediate carbocations, or do enzymes substantially reshape these constraints to produce end-products contributing to host fitness?

To investigate these questions, we created a series of model decalin carbocation systems and used density functional theory (DFT) to interrogate the conformers and transition state connectivities of the eudesm-5-yl carbocation leading to three major skeletal classes. We then applied theoretical morphology, an analytical methodology of paleobiology,[30] to analyze and interpret our computational results to classify and rationalize the distribution of terpene skeletal forms. Specifically, we defined a set of minimal geometric parameters of the eudesm-5-yl carbocation to construct a "morphospace" of possible conformations. We subsequently populated this morphospace with energetically

viable conformers from our DFT calculations to relate molecular geometries to cyclization paths. On the basis of this rigorous QM treatment, we were able to calculate the gas phase, theoretically expected frequency of eremophilane, spirovetivane, and spirojatamane skeletal forms. Finally, a comparison of theoretical to currently documented abundances revealed a perplexing disparity, indicating that the natural chemical abundance is skewed from the expected thermodynamic distribution. This divergence strongly suggests the adaptive significance of enzymatic traits underlying terpene skeletal forms in nature.

METHODS

Computational Details. All calculations were performed with Gaussian 98W[31] and the density functional method using B3LYP, Becke's three-parameter hybrid method[32] with the Lee−Yang−Parr correlation functional[33] and the 6-31G* basis set.[34] All stationary points were confirmed with second derivative calculations. Energies reported here include zero-point energy corrections calculated with unscaled B3LYP/6-31G* frequencies obtained analytically with G98W. Free energies of all C15 conformers were obtained from the output of Gaussian 98W second derivative (FREQ) calculations. Intrinsic reaction coordinate calculations[35,36] were used to determine reaction pathways. Additional details can be found in the Supporting Information.

Theoretical Yield of C-15 Products. We computed the gas-phase percent yield of all of the products found using the computed DFT energies of the reacting conformers and their transition structures. For the computations to be valid, it must be assumed that the equilibrium of all of the conformers is established at a much faster rate than the rates of reaction of the individual conformers. The concentrations of the reacting conformers are proportional to the mole fraction of each conformer present at equilibrium. Using the computed DFT free energies of the conformers, one can calculate equilibrium constants (K_i) connecting the conformers and from these the mole fractions (N_m) of each conformer present at equilibrium as developed earlier by Hur and colleagues.[37] A general formula for multiple conformers in equilibrium can easily be derived here for n conformers with $n - 1$ number of conformational equilibria:

$$N_m(n) = \frac{B_m}{1 + A(n)} \text{ for } m = 1, \ldots n$$

where $B_1 = 1$ and for $m > 1$:

$$B_m = \prod_{i=1}^{m-1} K_i$$

and

$$A(n) = \sum_{j=1}^{n-1} \prod_{i=1}^{j} K_i$$

Using the Arrhenius equation, we calculated the reaction rates (k_{ij}) from the computed activation energies for each reaction. Finally, we calculated the percentage (P_{ij}) of each product using the above computed mole fractions (note that the total number of reactions is greater than the number of conformers, because many of the conformers undergo more than one reaction):

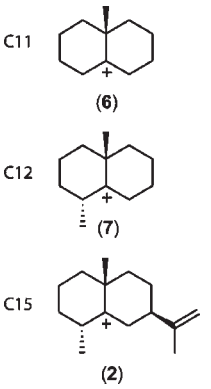
$$P_{ij} = \frac{N_i}{D} k_{ij} \times 100\%$$

where

$$D = \sum_i (N_i \sum_j^{r_i} k_{ij})$$

for $j = 1, 1 \ldots r_i$ for each conformer i, and N_i and k_{ij} are defined as above.

Scheme 2. Decalin Carbocation Model Systems[a]



[a] A series of model cyclic carbocation systems were constructed to evaluate the effect of ring substituents on the distribution of energetically favorable ring conformations.

RESULTS AND DISCUSSION

Constructing Carbocation Model Systems. To define the physical constraints governing the reactivity of the eudesm-5-yl carbocation (**2**), we considered a series of carbocation model systems (**2**, **6**, and **7**, Scheme 2). In earlier work, we calculated a low-energy conformer of **2** with transition state connectivity to methyl and methylene migrations leading to eremophilane and spirovetivane products, respectively.[29] In the current study, we extended our investigations to include the spirojatamane pathway (Figure 1, path a) along with two simplified models designed to probe the contribution of ring substituents to conformation and rearrangement pathways (Scheme 2). Specifically, we considered the simplest bicyclic structure, a methyldecalin C11 system (**6**), an achiral model, which retains only the potentially migrating methyl and methylene groups. We then added the C15 methyl group to produce the C12 system (**7**), while further addition of the isopropenyl group restores the C15 eudesm-5-yl carbocation system (**2**).

Computational Search of Conformational Space. Model systems in hand, we conducted a rigorous DFT computational search for conformers starting with the C12 system (**7**) using various combinations of boat and chair conformations of the decalin ring system. These efforts yielded 11 conformers (Figure 2a), which seeded further computational searches for the C11 (Figure 2b) and C15 (Figure 2c) systems. In the case of the C15 system (**2**), we also explored three possible rotamers of the isopropenyl tail, resulting in 31 of 33 potential conformers. Steric effects stemming from presence of the isopropenyl tail ultimately eliminated two possible conformers. Last, the conformational analyses of the C11 system produced 10 distinct conformers. The 31 C15 conformers were energetically within 10 kcal/mol of the most stable C15 conformer (1J-a). Therefore, our computational search identified diverse conformers of varying stabilities. All results are summarized in Figure 2 and Table 1.

Constructing and Populating a Morphospace. To globally analyze the conformational space occupied by the carbocation model systems, we employed a theoretical morphologic analysis. To begin, we first defined a minimal set of geometric parameters to describe the conformation and reactivity of the decalin ring; a pair of dihedral angles (dh1 and dh2) describe the migrating

methyl/methylene bonds and therefore directly relate transition state geometry to the potential cyclization pathways (Figure 1). We then retrieved these dihedral angles from the calculated conformers (Table 1) and plotted them against each other to populate a "morphospace". Visual analysis of the resulting plot revealed a scatter of conformers along a diagonal we term the dihedral axis, with two symmetrically opposed regions devoid of conformers for the C11 and C12 systems (Figure 3a). Physically, certain combinations of dihedral angles introduce significant torsional strain in the prescribed conformers, thereby raising the energy of such conformations creating substantial reaction barriers, an observation not previously noted. This region spans a 20 by 20° dihedral angle sweep (dh1/dh2) in conformational space, situated symmetrically at +40° to +60°/+5° to −15° and −40° to −60°/−5° to +15°.

Interestingly, the C15 system populates one of these barren regions (Figure 3b). To computationally probe this unexpected population of C15 conformers, we removed the isopropenyl moiety from conformers **1G-b** and **1H-a**, two C15 representatives from the barren region, and performed a DFT geometry optimization to locate energy minima. Each starting carbocation underwent "conformational readjustment" moving to a lower energy conformation **1H**, a C12 conformer, outside of the barren region (Table 1). Thus, while the isopropenyl tail of the C15 system introduces significant steric crowding depending on its rotation, this is offset by increased torsional strain, as exemplified by the **1G** conformers (Figure 3c). Therefore, direct comparison between model systems sheds light on the contribution of ring substituents to expanding accessible conformational space in the C15 decalin carbocation ring system.

Given the large collection of possible conformers, which ones are most relevant to subsequent cyclization reactions of the C15 system? The simplest explanation is that the lowest energy conformers will dominate, particularly if the system is under equilibrium. Therefore, we calculated the distribution of all conformers at equilibrium, having found minima on the potential energy surface. For this, we used free energies of the conformers, assuming they are in equilibrium with one another at 298.13 K (as detailed in Methods). To visualize this, we scaled the data points in our morphospace according to equilibrium abundance (Supporting Information Table 1). As expected, we observed a skewed distribution with many conformers assuming minimal population distributions

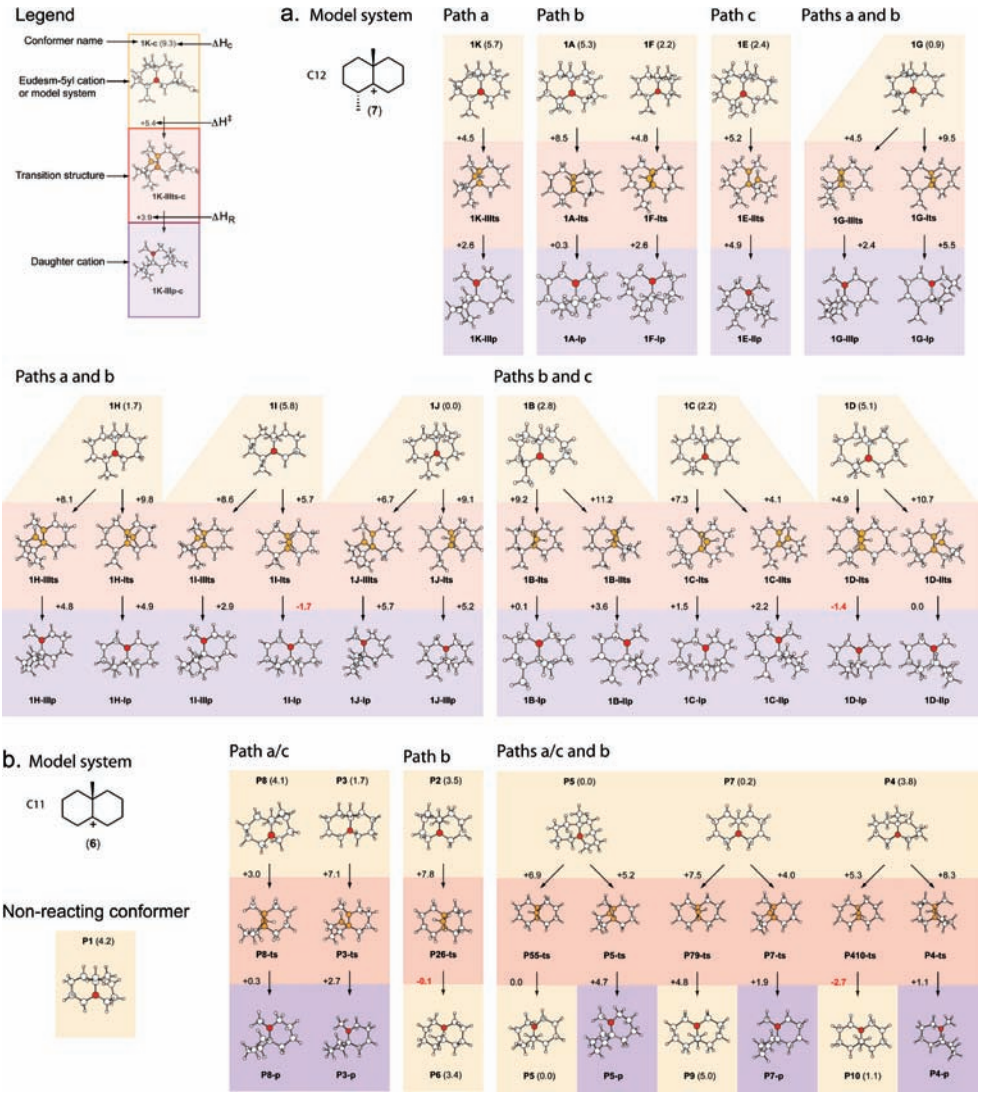


Figure 2. Continued

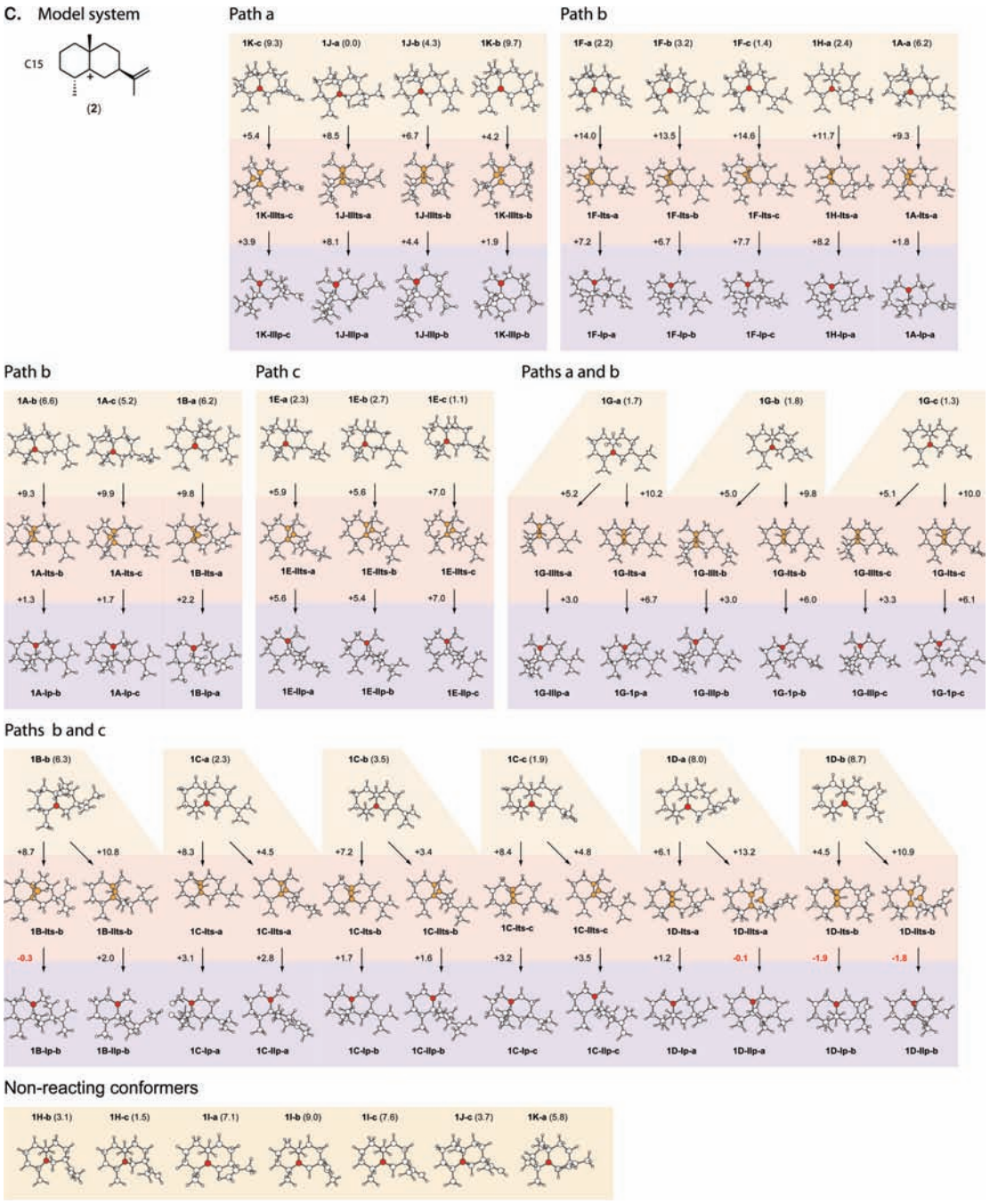


Figure 2. Conformers, transition structures, and products of decalin carbocation model system. Results of QM calculations for (a) C12 (**7**), (b) C11 (**6**), and (c) C15 (**2**) decalin models systems (Scheme 2) are illustrated as ball and stick models. A legend (inset) describes the arrangement of reactant conformer, transition structure, and product cations and the calculated energy of conformers (ΔH_c), activation energy ($\Delta H^{\ddagger}$), and energy of reaction (ΔH_R). Conformers are grouped according to cyclization pathways discovered by internal reaction coordinate calculations. Pathways correspond to Wagner−Meerwein rearrangements described in Figure 1. The red atoms in reactants and products indicate the location of positive charge, while gold atoms in the transition structures indicate the three allyl cation carbon atoms directly involved in the Wagner−Meerwein shifts. For some conformers, no transition structures were discovered. For the achiral C11 (**6**) system in panel b, some product cations correspond to reactant conformers found in initial calculations, as indicated by the shading scheme (inset legend).

(Figure 3d). These sparsely populated conformational spaces are due to the large energy differences between conformers, with the most stable conformer **1J-a** dominating the landscape. Interestingly, conformers residing in the barren region are reasonably stable relative to others, indicating their energetic viability and suggesting potential relevance to cyclization (described below).

Conformation to Pathways. To relate conformations to accessible cyclization pathways and hence products (Figure 1), we conducted a computational search for transition structures leading to the three potential Wagner−Meerwein shifts for each conformer. Once a transition structure was located, an internal reaction coordinate (IRC) calculation was performed, first to

dx.doi.org/10.1021/ja203342p |*J. Am. Chem. Soc.* 2011, 133, 12632–12641

Table 1. Conformations, Energy, and Reaction Pathways for the Carbocation Systems[a]

							Wagner–Meerwein shift		
system	cation	ΔH_c	$\Delta H^{\ddagger}$	ΔH_R	dh1	dh2	*a*	*b*	*c*
C15 (**2**)	1K-c	9.3	5.4	3.9	−22.0	−73.0	+		
	1J-c	3.7	NR	NR	−21.7	−70.4			
	1K-a	5.8	NR	NR	−20.3	−71.0			
	1J-a	0.0	8.5	8.1	−19.7	−68.9	+		
	1J-b	4.3	6.7	4.4	−16.5	−67.7	+		
	1K-b	9.7	4.2	1.9	−16.0	−70.0	+		
	1H-b	3.1	NR	NR	−13.4	−67.2			
	1H-c	1.5	NR	NR	−9.7	−63.1			
	1F-a	2.2	14.0	7.2	−7.3	−61.9		+	
	1G-b	1.8	5.0	3.0	1.0	−58.6	+		
		1.8	9.8	6.0	1.0	−58.6		+	
	1H-a	2.4	11.7	8.2	1.8	−52.9		+	
	1I-c	7.6	NR	NR	7.1	−46.0			
	1F-c	1.4	14.6	7.7	7.6	−45.3		+	
	1I-b	9.0	NR	NR	11.6	−40.4			
	1I-a	7.1	NR	NR	14.0	−37.8			
	1F-b	3.2	13.5	6.7	17.8	−38.5		+	
	1G-a	1.7	5.2	3.0	20.4	−41.0	+		
		1.7	10.2	6.7	20.4	−41.0		+	
	1G-c	1.3	5.1	3.3	21.5	−40.4	+		
		1.3	10.0	6.1	21.5	−40.4		+	
	1A-c	5.2	9.9	1.7	21.7	−29.2		+	
	1B-a	6.2	9.8	2.2	22.3	−24.7		+	
	1A-a	6.2	9.3	1.8	22.6	−29.1		+	
	1A-b	6.6	9.3	1.3	24.6	−27.7		+	
	1D-a	8.0	6.1	1.2	31.6	−21.6		+	
		8.0	13.2	−0.1	31.6	−21.6			+
	1B-b	6.3	8.7	−0.3	32.8	−18.1		+	
		6.3	10.8	2.0	32.8	−18.1			+
	1C-b	3.5	7.2	1.7	34.9	−20.3		+	
		3.5	3.4	1.6	34.9	−20.3			+
	1C-a	2.3	8.3	3.1	36.0	−19.9		+	
		2.3	4.5	2.8	36.0	−19.9			+
	1C-c	1.9	8.4	3.2	36.4	−19.1		+	
		1.9	4.8	3.5	36.4	−19.1			+
	1D-b	8.7	4.5	−1.9	36.6	−16.8		+	
		8.7	10.9	−1.8	36.6	−16.8			+
	1E-a	2.3	5.9	5.6	61.9	5.8			+
	1E-c	1.1	7.0	7.0	62.7	7.8			+
	1E-b	2.7	5.6	5.4	64.0	8.3			+
C12 (**7**)	1J	0	9.1	5.2	−14.1	−66.9		+	
	1J	0	6.7	5.7	−14.1	−66.9	+		
	1K	5.7	4.5	2.6	−13.2	−67.7	+		
	1H	1.7	9.8	4.9	−10.5	−65.8		+	
	1H	1.7	8.1	4.8	−10.5	−65.8	+		
	1F	2.2	4.8	2.6	−10	−65.7		+	
	1G	0.9	4.5	2.4	15	−46.3	+		
	1G	0.9	9.5	5.5	15	−46.3		+	
	1I	5.8	5.7	−1.7	18.9	−33.6		+	
	1I	5.8	8.6	2.9	18.9	−33.6	+		
	1A	5.3	8.5	0.3	23.2	−28.4		+	
	1C	2.2	7.3	1.5	34.6	−20.8		+	
	1C	2.2	4.1	2.2	34.6	−20.8			+
	1B	2.8	9.2	0.1	35.1	−16.6		+	
	1B	2.8	11.2	3.6	35.1	−16.6			+
	1D	5.1	4.9	−1.4	35.9	−16.5		+	
	1D	5.1	10.7	0	35.9	−16.5			+
	1E	2.4	5.2	4.9	64.4	8.3			+
C11 (**6**)	P8	4.1	3	0.3	−11.9	−65.5	+		
	P6	3.4	7.9	0.1	19	−34.5		+	
	P1	4.2	NR	NR	21.3	−28.4			
	P9	5	2.7	5	21.8	−21.8		+	
	P7	0.2	7.5	4.8	30.2	−30.1		+	
		0.2	4	1.9	30.2	−30.1	+		
	P10	1.1	8	1.1	33.6	−20.5		+	
	P2	3.5	7.8	−0.1	35.3	−15		+	
	P4	3.8	5.3	−2.7	35.6	−15.4		+	
		3.8	8.3	1.1	35.6	−15.4	+		
	P3	1.7	7.1	2.7	64	10.1	+		
	P5	0	6.9	0	65.4	13.1		+	
		0	5.2	4.7	65.4	13.1	+		

[a] ΔH_c: Energy (in kcal/mol) of the conformer relative to **1j-a, 1J**, and **P5** for the C15, C12, and C11 systems, respectively. $\Delta H^{\ddagger}$: Activation energy (reactant to transition structure in kcal/mol). ΔH_R: Energy of reaction (reactant to product) in kcal/mol. dh1 and dh2: Dihedral angles as defined in Figure 1.

ensure linkage to the conformer under investigation and second to find the product carbocation. The results are summarized in Figure 2 and Table 1. In almost all cases, these migration reactions are endothermic, suggesting a role for the enzyme in controlling this aspect of the cyclization reaction. Annotating the morphospace with transition state connectivity illustrates how each Wagner–Meerwein shift distributes along the axis of accessible conformations (Figure 4).

Analysis of the simple C11 model system (**6**) reveals fundamental geometrical features that relate to cyclization reactivity. This system is unique in having a plane of symmetry passing through the methyl carbon and the two bridgehead carbons and is therefore achiral (dashed line, Figure 4a). Interestingly, two conformers (**P7** and **P9**) lie on this plane of symmetry and are connected via methyl migrations. For example, conformer **P7** has equal dihedral angles 1 and 2 (+30°, −30°) and parents a methyl migration leading to **P9** and a distinct methylene migration pathway leading to a spiro cation (Figure 2b). Dual transition state connectivity turns out to be common in the more complex C12 and C15 systems (described below); however, the methylene migrations arising from **P7** are indistinguishable as "enantiomeric pathways" given the symmetry. Conformer **P9** also has equal dihedral angles (+22°, −22°), but gives rise only to the methyl migration pathway. Therefore, it appears that a 30° dihedral angle threshold is required for methylene migrations to occur. This means that the migrating bond in the "reactant" carbocation only needs to be within 60° of perpendicularity to the plane of the carbocation center for the reaction to proceed,

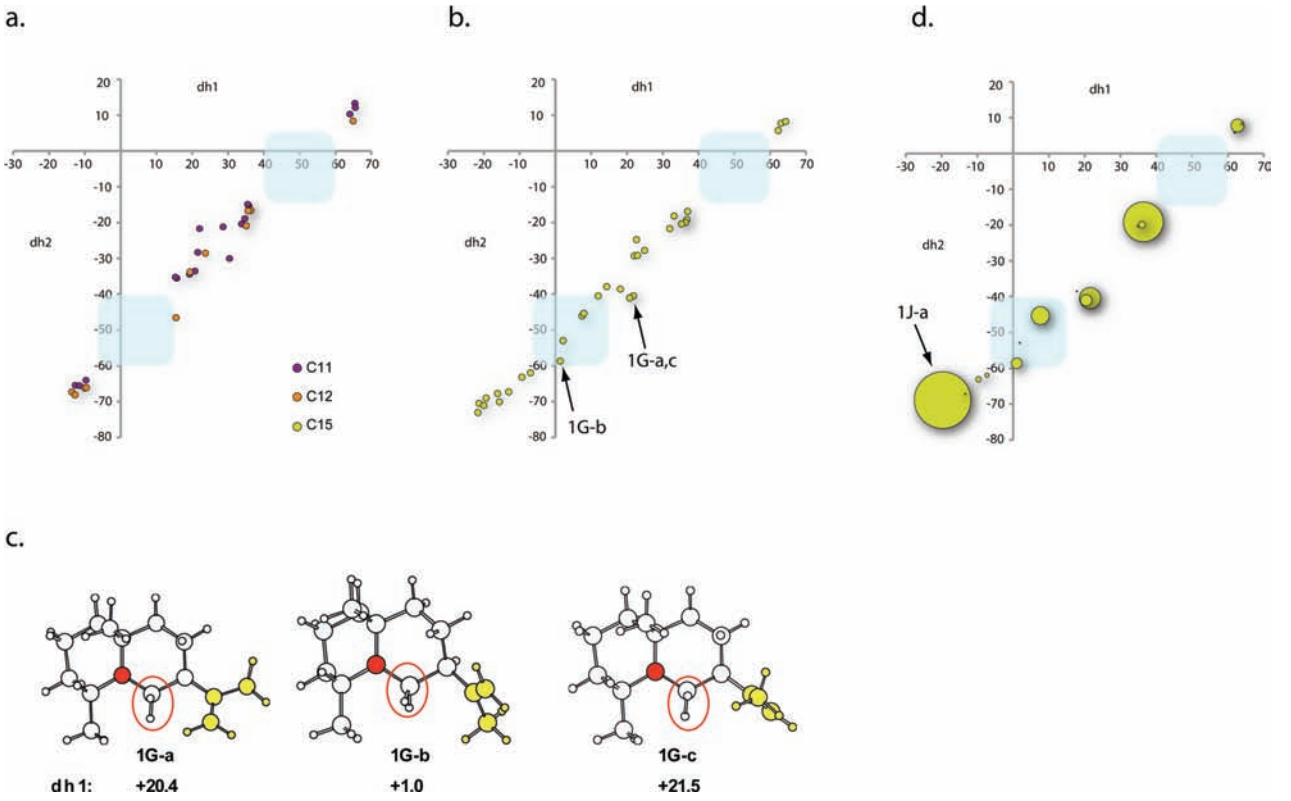


Figure 3. Model carbocation systems and the distribution of low-energy conformers. (a) Scatter plot of dihedral angles 1 and 2 from energetically favorable conformations of the C11, C12, and (b) C15 systems derived from DFT calculations (Figure 2, Table 1). Shading indicates barren regions in the plot, two for the C11 and C12 conformers, one of which is shared with the C15 conformers to represent the barren region where no low-energy conformers were found. (c) A representative conformer (**1G-b** in Figure 2) of the C15 system (**2**) from the barren region of the dihedral plot is illustrated as viewed along the carbocation plane down the C2/C7 bond axis (top) or by 90° rotation. (d) Conformers of the C15 system where sphere size was scaled according to equilibrium concentrations (Supporting Information Table 1).

a previously unreported observation to our knowledge, and a unique contribution from the current work.

Extending our analysis to more complex systems exposes added constraints on reactivity imposed by ring substituents. Introduction of the methyl group in the C12 system (**7**) breaks the internal mirror plane symmetry; no such conformer akin to **P7** or **P9** was discovered for the C12 and the more highly substituted C15 systems. However, the 30° dihedral angle forms a dividing line separating methylene migration pathways leading to either spirovetivane or spirojatamane products, which minimally require a dh1 ≥ +30° or dh2 ≥ −30°, respectively (Figure 4b and c). Although many conformers display dual transition state connectivity, several clusters of conformers share connectivity to single cyclization pathways. For example, spirane-exclusive pathways occupy the poles at extreme ends of the plot as expected given the opposing geometrical requirements for their respective Wagner−Meerwein shifts (Figure 1, paths a and c). In the center of the plot is a cluster exclusively linked to methyl migrations, while a sparser collection extends clear into the spirojatamane end of the dihedral axis. Interestingly, there is a tight cluster with dual spirovetivane and eremophilane connectivity, in contrast to the scattered spread of conformers connecting the other set of dual pathways. In some instances, no transition state could be found linking a carbocation to any transition state, particularly for the C15 system. Hence, these conformers represent "dead ends" due to steric blockage from the methyl and/or isopropenyl group. Thus, while the ring substituents expanded accessible conformational space, the associated steric effects block rearrangements in certain contexts.

To summarize the pathway connectivity results, we related the dihedral scatter plot back to our Newman projection model of eudesm-5-yl carbocation system (**2**) (Figure 4d). The methyl group sweeps through a 45° arc in conformers where the methyl migration pathway is physically viable, whereas the methylene migrations have a narrower window of 35° by comparison. In all cases, only a narrow arc is exclusive to a given pathway. Therefore, the enzyme active site may be sensitively tuned to filter out closely related conformers given the overlap in transition state connectivity.

Theoretical Abundance of Sesquiterpene Skeletal Types. To determine the expected frequency of sesquiterpene skeletal types based on conformation and reactivity of the eudesm-5-yl carbocation (**2**), we calculated the theoretical abundance. Using energies of activation (Figure 2c and Table 1), we calculated the first-order rate constants for each Wagner−Meerwein rearrangement. Weighting each rate constant with mole fractions from equilibrium calculations (Supporting Information Table 1), we obtained the percentages of the three main pathways, and hence the theoretical final product outcomes (Table 2). Strikingly, the gas-phase calculations indicate the two methylene shift pathways should dominate, with paths a (spirojatamane) and c (spirovetivane) constituting 56.3% and 43.6% product, respectively. By contrast, the methyl migration, which can occur over the largest swath of conformational space (Figure 4d), only contributes 0.1% to product outcomes. These predictions are in stark contrast to

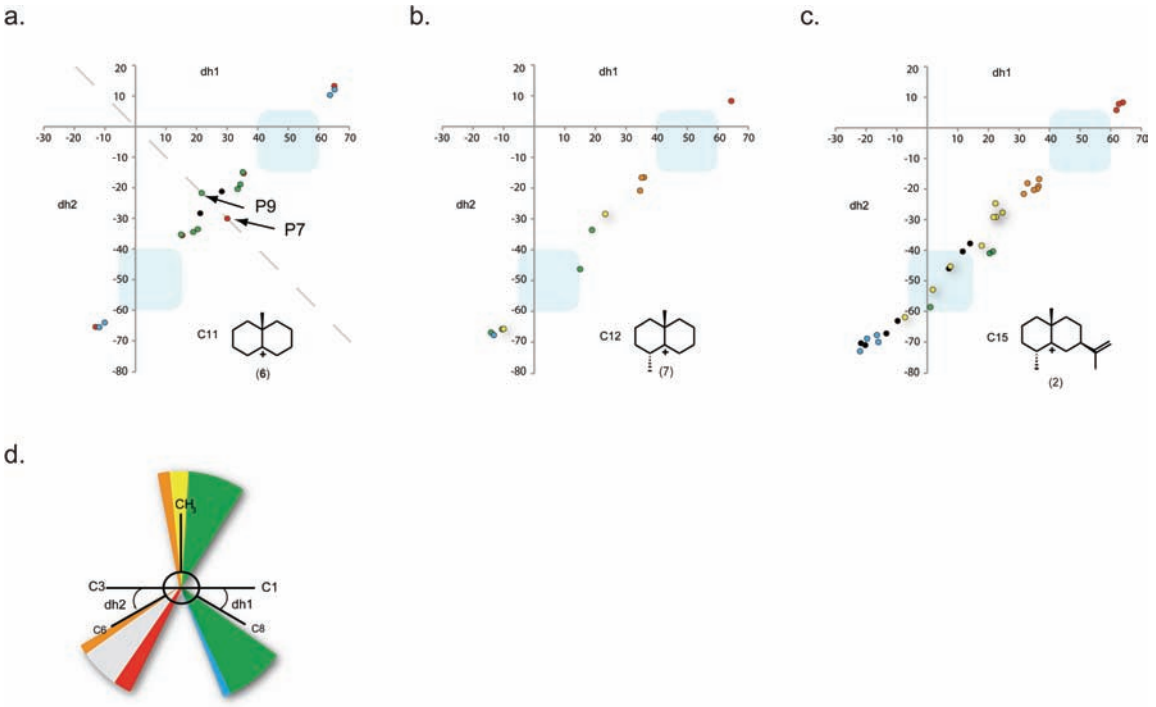


Figure 4. Geometric parameters and cyclization pathways. (a−c) The transition state connectivity of carbocations from the C12 and C15 systems is annotated on the dihedral angle scatter plot according to color, path a (blue), path b (yellow), path c (red), path a/b (green), path b/c (orange), and no reaction (black). For the C11 system, exclusively spiro (blue), eremophyl (green), or both pathways (red) are indicated, and the dashed line denotes the mirror-plane symmetry of this system. (d) Newman projection (Figure 1) annotated to show the dihedral sweep of each potentially migrating carbon.

Table 2. Theoretical Distribution of Cyclization Products Arising from the Eudesm-5-yl Carbocation System (2)[a]

conformer	% abundance	W−M shift	product
1C-a	12.7	**c**	spirovetivane
1C-b	13.8	**c**	spirovetivane
1C-c	14.7	**c**	spirovetivane
1E-a	0.9	**c**	spirovetivane
1E-b	0.8	**c**	spirovetivane
1E-c	0.7	**c**	spirovetivane
1G-a	12.4	**a**	spirojatamane
1G-b	16.3	**a**	spirojatamane
1G-c	27	**a**	spirojatamane
1J-a	0.5	**a**	spirojatamane

[a] Values were calculated as described in methods and listed in the Supporting Information.

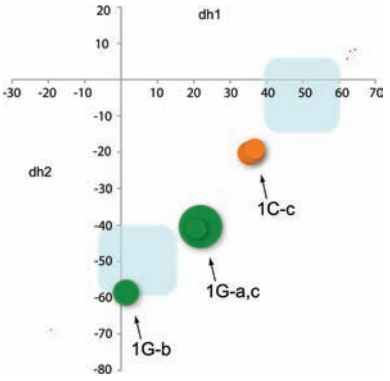


Figure 5. Conformers contributing to the theoretical abundance of cyclization products. Conformers along the dihedral axis of the C15 morphospace are scaled according to their contribution to the overall abundance of terpene skeletal types as listed in Table 1 (color scheme as in Figure 4). Shading denotes a barren area.

the documented abundance in the *Solanaceae* (albeit a currently sparse sample), which produce spirovetivane and eremophilane skeletal classes in abundance, while the spirojatame skeleton is absent and only rarely observed in other organisms.

To visually analyze the theoretical abundance results in morphospace, we scaled the spheres on the dihedral axis according to their contribution to predicted product outcomes (Figure 5). Our results indicate that thermodynamically, only a few conformers are relevant to final product outcomes, all of which have dual transition state connectivity, including **1G-b**, which is from the barren region. Interestingly, while **1J-a** is the most energetically favorable conformer (Figure 3d) and has exclusive linkage to spirojatamane (Figure 2c), this conformer contributes little to the products.

Comparison of physical models of reactant, transition state, and product carbocations along the dominant reaction coordinates reveals a counterintuitive result. Given the endothermic nature of these rearrangements, we observe the structure of transition states most closely resembles products according to Hammond's postulate, as expected (Figure 6, Table 1). Strikingly, methylene migration pathways, while lower in energy, exhibit considerably greater structural change during these rearrangements (Figure 6). Conversely, methyl migrations are high-energy pathways relative to the methylene migrations, although they exhibit substantially fewer structural rearrangements in traversing the transition state to product carbocations.

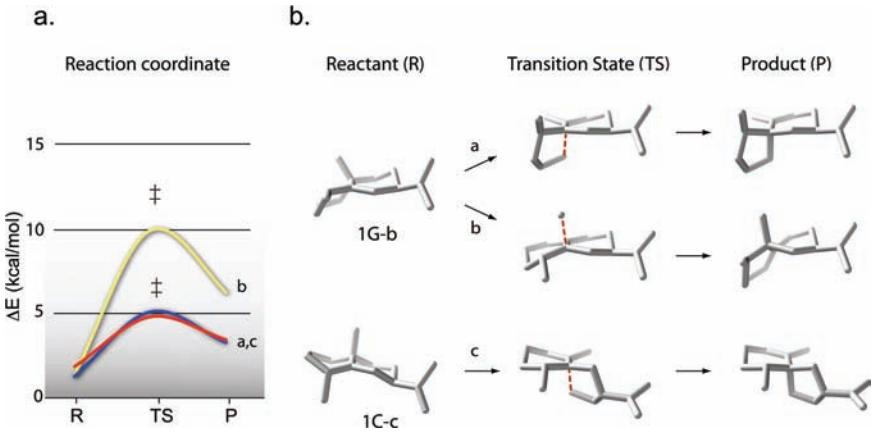


Figure 6. Structural transformations along the reaction coordinate. (a) A reaction coordinate diagram depicting the energy change along representative pathways from low energy conformer reactant (R), transition state (TS), to product (P) carbocations where paths a and b derive from **1G-b** and path c from **1C-c** reactant cations, respectively. (b) Structural depiction of representative pathways from panel a where dashed red lines indicate newly forming bonds in each respective transition state. Hydrogen atoms were removed for clarity.

Steric effects from the C15 methyl and isopropenyl groups should elevate the methyl migration energy, and indeed should block this pathway in some cases. However, even in the absence of ring substituents, as in the minimal C11 system (**6**), where the methyl group is perfectly aligned for migration in **P7**, this pathway is still considerably higher in energy (3.5 kcal/mol) than the alternative methylene shift path.

■ CONCLUSIONS

This Article critically examines how the conformational state of a pivotal carbocation intermediate relates to branching cyclization pathways and ultimately the sesquiterpene skeletal forms at the core of diverse terpene structures in nature. Borrowing analytical tools from paleobiology, the current work presents a novel approach to analyzing and interpreting a QM analysis of carbocations; by defining minimal geometric parameters that describe the carbocation structure, we were able to globally analyze and compare model systems in a theoretical morphospace to reveal a region of unexpectedly high activation barriers. This led to further insights into the contribution of ring substituents to conformation and reactivity of the eudesm-5-yl carbocation, which ultimately allow entry into this region and further give rise to energetically favorable transition states. Furthermore, our calculations reveal a 30° rule, which defines a geometrical requirement in the migrating bond of the reactant carbocation for viable transition state connectivity and hence rearrangement reactions to proceed.

A major caveat remains; these are gas-phase calculations whereas cyclization pathways ultimately transpire in the confines of a dynamic enzyme active site, restricted by geometry, stereoelectronic effects, bound solvent, and electrostatics, most likely modulated by varying distributions of enzyme conformational states. Furthermore, product outcomes may be dictated by kinetic control as previously suggested for sesquiterpenes from trichodiene synthase[38] and for monoterpene products given the wide-ranging heats of formation from the α-terpinyl cation.[39] Despite these limitations, QM calculations currently provide the only means of reconstructing the geometry and intrinsic reactivity of these species. The theoretical morphologic approach used here provides a useful framework for interpreting the conformational and energetic landscape that underpins intrinsic reactivity. Controlling this landscape is the central challenge of the enzyme and is clearly under evolutionary restraints. Consider the TEAS product spectrum that contains both spirovetivane and spirojatamane products in minor amounts. These products originate from distinct conformer origins of the eudesm-5-yl cation that populate the active site, yet are effectively filtered out by the enzyme most of the time. However, examination of transition state models in the static active site crystal structure of TEAS reveals no obvious steric or electronic rationale for selection of the eremophilane or exclusion of the spiro cyclization pathways. Although accurate conformer filtering may be important, most conformers are linked to two separate pathways, as illustrated from the annotated morphospace (Figure 4). Therefore, we speculate the enzyme may have evolved to recognize features arising from the sequence of structural changes in the cyclizing species as another level of discrimination, as these changes may be quite substantial (Figure 6).

Quantum mechanics has enabled us to precisely calculate the relative energies and intrinsic reactivities of conformations of the eudesm-5-yl carbocation. Our effort uniquely contributes to a growing body of work aimed at producing a detailed physical picture and deeper insights into terpene cyclization. For example, the calculation of polycyclization pathways of the farnesyl cation leading to the multicyclic sesquiterpene pentalenene was found to involve unusual and unexpected intermediates,[40] and quantum chemical calculations produced a physical model relating several sesquiterpene families to bisabolyl cation cyclization mechanisms.[41] Extensive DFT calculations on the cyclization of squalene to lanosterol and hopenes have shown these cyclizations to be not only by and large concerted in nature,[42−44] but also to avoid the previously proposed anti-Markovnikov ring expansion in the formation of the C and D rings in these cyclizations.[45] More recently, DFT calculations coupled with X-ray crystallography and organic chemistry enabled the structural elucidation of the cisoid cyclization pathway in TEAS.[46]

While the QM work presented here is a central component of a model describing the physical constraints underlying terpene chemical diversity, further expansion with computational approaches like QM/MM will be invaluable in probing the potential influence of macroscopic vibrational modes of the

protein with the cyclizing carbocation. Coupling these efforts with experimental approaches using NMR and isotopic labeling, integrated with crystallographic information, will likely lead to the proposal of a new theoretical abundance of products based on inclusion of additional physical constraints introduced by the enzyme. Finally, existing reports indicate that solanaceous plants produce primarily skeletons from paths b and c (5-EA and PSD) with no report of spirojatamane derivatives found in this family; a comprehensive cataloguing of the natural abundance of terpene skeletal classes and characterization of their cognate terpene cyclases will be important future work as we endeavor to piece together the distribution and adaptive significance of terpene skeletal forms in nature.

■ ASSOCIATED CONTENT

Supporting Information. One table and for all structures considered G98W output showing Cartesian coordinates, lowest three vibrational frequencies, and energies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors
paul.omaille@bbsrc.ac.uk; b.andes.hess@vanderbilt.edu

■ ACKNOWLEDGMENT

We would like to thank George McGhee for insightful discussions and comments on the manuscript, and C. Crenshaw and A. Osbourn for critical evaluation of the manuscript.

■ REFERENCES

(1) Connolly, J. D.; Hill, R. A. *Dictionary of Terpenoids*; Chapman and Hall: London, 1991.
(2) Akiyama, K.; Matsuzaki, K.; Hayashi, H. *Nature* **2005**, *435*, 824.
(3) Grayer, R. J.; Kokubun, T. *Phytochemistry* **2001**, *56*, 253.
(4) Harborne, J. B. *Biochem. Syst. Ecol.* **1999**, *27*, 335.
(5) Pedras, M. S. C.; Okanga, F. I.; Zaharia, I. L.; Khan, A. Q. *Phytochemistry* **2000**, *53*, 161.
(6) Pichersky, E.; Gershenzon, J. *Curr. Opin. Plant Biol.* **2002**, *5*, 237.
(7) Reinhard, J.; Srinivasan, M. V.; Zhang, S. W. *Nature* **2004**, *427*, 411.
(8) Di Santo, R. *Nat. Prod. Rep.* **2010**, *27*, 1084.
(9) Krings, U.; Berger, R. G. *Nat. Prod. Commun.* **2010**, *5*, 1507.
(10) Degenhardt, J.; Kollner, T. G.; Gershenzon, J. *Phytochemistry* **2009**, *70*, 1621.
(11) Austin, M. B.; O'Maille, P. E.; Noel, J. P. *Nat. Chem. Biol.* **2008**, *4*, 217.
(12) Croteau, R. B.; Wheeler, C. J.; Cane, D. E.; Ebert, R.; Ha, H. J. *Biochemistry* **1987**, *26*, 5383.
(13) Schenk, D. J.; Starks, C. M.; Manna, K. R.; Chappell, J.; Noel, J. P.; Coates, R. M. *Arch. Biochem. Biophys.* **2006**, *448*, 31.
(14) Wagschal, K.; Savage, T. J.; Croteau, R. *Tetrahedron* **1991**, *47*, 5933.
(15) Tantillo, D. J. *Nat. Prod. Rep.* **2011**, *28*, 1035.
(16) Cane, D. E. *Acc. Chem. Res.* **1985**, *18*, 220.
(17) Mercke, P.; Crock, J.; Croteau, R.; Brodelius, P. E. *Arch. Biochem. Biophys.* **1999**, *369*, 213.
(18) O'Maille, P. E.; Chappell, J.; Noel, J. P. *Arch. Biochem. Biophys.* **2006**, *448*, 73.
(19) Seaman, F. C. *Bot. Rev.* **1982**, *48*, 121.
(20) Zidorn, C. *Phytochemistry* **2008**, *69*, 2270.
(21) Spring, O. *Recent Adv. Phyochem.* **1991**, *25*, 319.
(22) Gordon, M.; Stoessl, A.; Stothers, J. B. *Can. J. Chem.* **1973**, *51*, 748.
(23) Burden, R. S.; Rowell, P. M.; Bailey, J. A.; Loeffler, R. S. T.; Kemp, M. S.; Brown, C. A. *Phytochemistry* **1985**, *24*, 2191.
(24) Metlitsk, L.; Ozeretsk, O.; Vulfson, N. S.; Chalova, L. I. *Dokl. Akad. Nauk SSSR* **1971**, *200*, 1470.
(25) Katsui, N.; Murai, A.; Takasugi, M.; Imaizumi, K.; Masamune, T.; Tomiyama, K. *Chem. Commun.* **1968**, 43.
(26) Fujimori, T.; Kasuga, R.; Kaneko, H.; Noguchi, M. *Phytochemistry* **1977**, *16*, 392.
(27) Eggers, M. D.; Sinnwell, V.; Stahl-Biskup, E. *Phytochemistry* **1999**, *51*, 987.
(28) Bagchi, A.; Oshima, Y.; Hikino, H. *Tetrahedron* **1990**, *46*, 1523.
(29) O'Maille, P. E.; Malone, A.; Dellas, N.; Hess, B. A., Jr.; Smentek, L.; Sheehan, I.; Greenhagen, B. T.; Chappell, J.; Manning, G.; Noel, J. P. *Nat. Chem. Biol.* **2008**, *4*, 617.
(30) McGhee, G. R. *Theoretical Morphology: The Concept and Its Applications*; Columbia University Press: New York, 1999.
(31) Frisch, M. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(33) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(34) Harihara, P.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(35) Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.
(36) Gonzalez, C.; Schlegel, H. B. *J. Phys. Chem.* **1990**, *94*, 5523.
(37) Hur, D.; Guven, A. *J. Mol. Struct. (THEOCHEM)* **2002**, *583*, 1.
(38) Vedula, L. S.; Rynkiewicz, M. J.; Pyun, H. J.; Coates, R. M.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2005**, *44*, 6153.
(39) Brandt, W.; Bräuer, L.; Günnewich, N.; Kufka, J.; Rausch, F.; Schulze, D.; Schulze, E.; Weber, R.; Zakharova, S.; Wessjohann, L. *Phytochemistry* **2009**, *70*, 1758.
(40) Gutta, P.; Tantillo, D. J. *J. Am. Chem. Soc.* **2006**, *128*, 6172.
(41) Hong, Y. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2009**, *131*, 7999.
(42) Gao, D. Q.; Pan, Y. K.; Byun, K.; Gao, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 4045.
(43) (a) Hess, B. A., Jr. *Org. Lett.* **2003**, *5*, 165. (b) Hess, B. A., Jr. *Collect. Czech. Chem. Commun.* **2003**, *68*, 202. (c) Hess, B. A., Jr.; Smentek, L. *Mol. Phys.* **2004**, *102*, 1201. (d) Hess, B. A., Jr. *Collect. Czech. Chem. Commun.* **2004**, *69*, 261. (e) Hess, B. A., Jr. *Eur. J. Org. Chem.* **2004**, 2239. (f) Hess, B. A., Jr.; Smentek, L. *Org. Lett.* **2004**, *6*, 1717. (g) Hess, B. A., Jr.; Smentek, L. *Collect. Czech. Chem. Commun.* **2008**, *73*, 786. (h) Smentek, L.; Hess, B. A., Jr. *J. Am. Chem. Soc.* **2010**, *132*, 17111.
(44) (a) Matsuda, S. P. T.; Wilson, W. K.; Xiong, Q. B. *Org. Biomol. Chem.* **2006**, *4*, 530. (b) Xiong, Q.; Rocco, F.; Wilson, W. K.; Xu, R.; Ceruti, M.; Matsuda, S. P. T. *J. Org. Chem.* **2005**, *70*, 5362.
(45) Hess, B. A., Jr. *J. Am. Chem. Soc.* **2002**, *124*, 10286.
(46) Noel, J. P.; Dellas, N.; Faraldos, J. A.; Zhao, M.; Hess, B. A., Jr.; Smentek, L.; Coates, R. M.; O'Maille, P. E. *ACS Chem. Biol.* **2010**, *5*, 377.