ORGANIC LETTERS

2007
Vol. 9, No. 11
2183−2186

Stereochemistry of Water Addition in Triterpene Synthesis: The Structure of Arabidiol

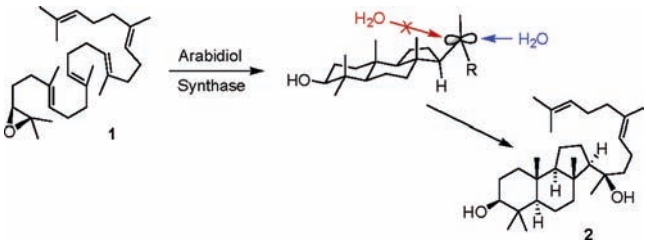
Mariya D. Kolesnikova,[†] Allie C. Obermeyer,[†] William K. Wilson,[‡] David A. Lynch,[‡] Quanbo Xiong,[‡] and Seiichi P. T. Matsuda*[,†,‡]

Department of Chemistry and Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005

matsuda@rice.edu

Received March 22, 2007

ABSTRACT



An oxidosqualene cyclase from *Arabidopsis thaliana* makes arabidiol, a tricyclic triterpene reported with indeterminate side-chain stereochemistry. We established the full structure of arabidiol through chemical degradation, NOE experiments, and molecular modeling. By examining the mechanistic constraints that govern water addition in triterpene synthesis, we further show how the stereochemistry of hydroxylation can generally be deduced a priori, why deprotonation is more common than hydroxylation, and why cyclases that perform hydroxylation also generate olefinic byproducts.

Triterpene synthases convert oxidosqualene (**1**) and squalene to carbocyclic skeletons through cation−π annulations and rearrangements that are usually quenched by deprotonation of the cationic intermediate to an olefin.[1] However, some enzymes add water to produce an alcohol instead.[1b] An example is At4g15340, *Arabidopsis thaliana* arabidiol synthase, which was recently reported to make exclusively a tricyclic diol (**2**) with indeterminate configuration at C14.[2] We now report the complete structure of **2**, which has the 14*R* configuration. This result prompted us to evaluate the mechanistic constraints that affect the orientation of cation hydroxylation in triterpene synthesis. Insights from this analysis explain the carbinol stereochemistry of known triterpene products of cation hydroxylation, their rarity in triterpene biosynthesis, and the likelihood of minor products.

At4g15340 cDNA was PCR-amplified from an *A. thaliana* cDNA pool and expressed in the yeast strains SMY8 and RXY6 from the vector pRS426GAL.[3] In vitro (RXY6) and in vivo (SMY8) experiments gave **2** as the dominant product, which was isolated by silica gel chromatography. 1D and 2D NMR spectra of **2** confirmed its tricyclic malabarica-17,21-diene-3β,14-diol structure except for the C14 configuration.

Establishing the C14 configuration of **2** was nontrivial. C14 is located in the side chain, where free rotation about C−C bonds generates conformational heterogeneity that complicates the interpretation of NOE results. To reduce the number of conformers, we degraded **2** to lactone **3** through

[†] Department of Chemistry.
[‡] Department of Biochemistry and Cell Biology.

(1) Reviews: (a) Wendt, K. U.; Schulz, G. E.; Corey, E. J.; Liu, D. R. *Angew. Chem., Int. Ed.* **2000**, *39*, 2812−2833. (b) Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261−290. (c) Abe, I.; Rohmer, M.; Prestwich, G. D. *Chem. Rev.* **1993**, *93*, 2189−2206. (d) Yoder, R. A.; Johnston, J. N. *Chem. Rev.* **2005**, *105*, 4730−4756.

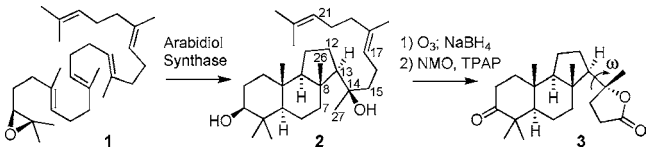
(2) Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835−2838.

(3) Hart, E. A.; Hua, L.; Darr, L. B.; Wilson, W. K.; Pang, J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1999**, *121*, 9887−9888.

10.1021/ol070709b CCC: $37.00
Published on Web 05/03/2007

a sequence of reactions that preserve the C14 configuration of **2** (Scheme 1).[4]

Scheme 1. Formation of **2** and Its Degradation to Lactone **3**



Lactone **3** contains a single freely rotatable dihedral angle ω. The Boltzmann distribution of staggered rotamers from B3PW91/6-311G(2d,p)//B3LYP/6-31G* energies[5] showed that the 14*R* epimer strongly favors the −gauche rotamer, whereas the 14*S* epimer adopts −gauche, +gauche, and anti conformations to comparable extents (Table 1). This distribution of rotamers was consistent with observed and calculated NMR shieldings, the root mean square (rms) deviations being lower for the rotamer distributions in Table 1 than for other combinations of rotamers (Table S4, Supporting Information).

Table 1. Elucidation of the C14 Configuration of **3**. Comparison of Observed NOE Enhancements against Interatomic Distances[a]

	14*R* epimer of **3**			14*S* epimer of **3**			
hydrogen pairs	g⁻ 80%	g⁺ 18%	a 2%	g⁻ 37%	g⁺ 33%	a 31%	obsd NOE
			distance (Å)				
27−12β		3.1	2.1		2.5	2.1	none
26−15*R*			2.8	2.3			none
7β−15*S*			2.4		2.0		none
27−7α	2.8			2.9			none
27−7β	2.2			2.2			strong
27−26	2.3		2.1	2.7		2.0	strong

[a] g⁻, g⁺, and a designate −gauche, +gauche, and anti rotamers about the H−C13−C14−C27 dihedral. Interatomic distances are given for the pairs of hydrogen atoms indicated; blank entries denote distances > 3.2 Å.

Results of NOESY and 1D NOE difference experiments for lactone **3** are summarized in Table 1. The absence of enhancements for hydrogen pairs 27−12β, 26−15*R*, and 7β−15*S* are all compatible with the 14*R* epimer (distances >2.8 Å for the major rotamers) but not the 14*S* epimer (distances <2.3 Å in a major rotamer for each of the 3 hydrogen pairs). The last three entries (27−7α, 27−7β, and 27−26) do not discriminate between the 14*R* and 14*S* epimers but illustrate that NOE enhancements were strong for pairs of hydrogens within 2.3 Å and essentially absent for pairs >2.8 Å apart. A comprehensive table of hydrogen pairs (Table S5, Supporting Information) further supported the 14*R* epimer. A similar set of calculations and NOESY data for diol **2** also strongly favored the 14*R* configuration (see the Supporting Information). Based on these two independent structure analyses, we conclude that arabidiol (**2**) is (13*R*,14*R*,17*E*)-malabarica-17,21-diene-3β,14-diol.

Our analysis of crude arabidiol synthase extracts from several in vitro and in vivo experiments showed many minor triterpene alcohols (Figure 1). Elucidating their structures is beyond the scope of this communication, but we describe the most abundant minor product **4**. Its mass spectrum resembled that of thalianol ((13*R*,14*R*,17*E*)-podioda-8,17,21-trien-3β-ol, **5**),[6] but its GC retention time was longer (Figure 1). 1D and 2D NMR experiments showed that **4** and **5** share identical overall connectivity and ring stereochemistry. The remaining stereocenter (C14) was established by the methodology described above for **2**. Thus, the Boltzmann distribution of conformers was determined for **4** and **5** and validated by NMR shielding calculations. Comparison with our NOESY results for both epimers showed **4** to be (13*R*,14*S*,17*E*)-podioda-8,17,21-trien-3β-ol, to which the trivial name 14-epithalianol was assigned. Our characterization of both C14 epimers of thalianol resulted in the

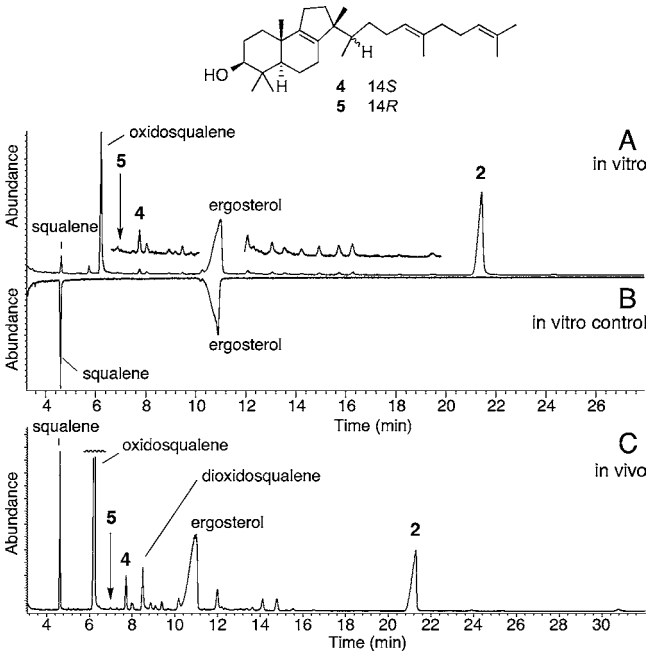


Figure 1. Total ion chromatogram from GC−MS of the crude products of arabidiol synthase (TMS derivatives; 260 °C oven temperature): in vitro reaction (A) with and (B) without oxidosqualene; (C) in vivo experiment. Note that **5** is absent in each chromatogram.

(4) Ozonolysis of **2** in CH_2Cl_2/CH_3OH (1:1) at −78 °C, followed by treatment with $NaBH_4$, gave a triol intermediate, which was oxidized to **3** with NMO, TPAP, and 4 Å molecular sieves in CH_2Cl_2 at room temperature.

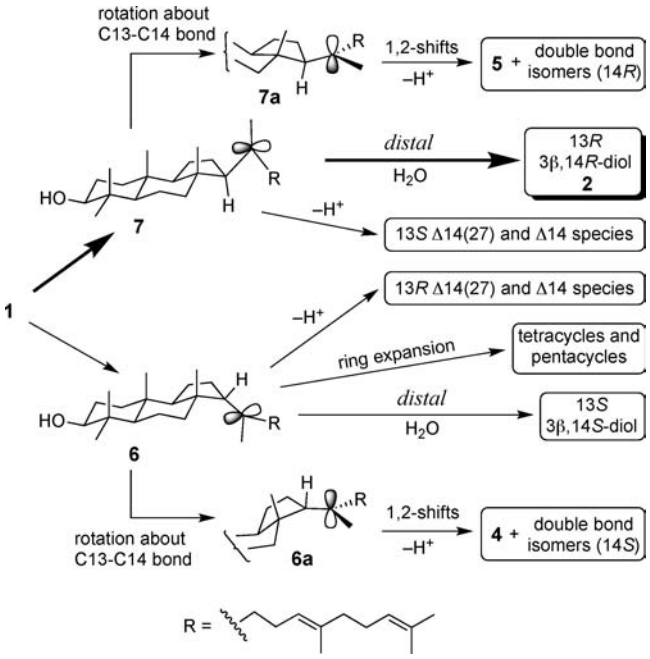
(5) Molecular modeling was done with Gaussian software: Frisch, M. et al. *Gaussian 03*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2005. The full reference is given in the Supporting Information.

(6) Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678−5679. The correct structure of **5** was shown, but the systematic name was given incorrectly as (3*S*,13*S*,14*R*)-malabarica-8,17,21-trien-3-ol.

assignment of several tricyclic triterpenes that have been reported with indeterminate configuration at C14 (see the Supporting Information). NMR and GC−MS analyses of the in vitro reaction products showed 14-epithalianol at 4−5% of the level of arabidiol.

Mechanistic origins of arabidiol, 14-epithalianol, and other minor products are suggested in Scheme 2. The typical

Scheme 2. Pathways Leading to **2** and Other Possible Products



tricyclic precursor of tetracyclic and pentacyclic triterpenes in plants is the 13*R* malabaricadienyl cation **6**. However, its 13*S* epimer **7** is clearly the precursor of arabidiol. Cation−π annulation forms **7** as a horizontal[7] cation, with its empty 2p orbital in the plane of the ring system and the 14-methyl anti to H13α. Conformer **7** can rotate about C13−C14 to make vertical[7] cation **7a**, but 90° rotation in the other direction[8] and 180° rotation[9] are precluded by the limited breadth of cyclase active site cavities.

To form arabidiol, a replenishable ordered water in the active site might attack the C14 cation of **7** on its proximal side (facing the cyclized core) or its distal side (facing the entrance channel of the active site) or attack **7a** from above or below. We have investigated these possibilities by adding dummy oxygen atoms to a B3LYP/6-31G* model[5] of **7**. To model proximal or distal attack, dummy atoms were placed 2.5 Å from C14 on the axis of the empty 2p orbital (Figure 2). For attack from above or below, dummy atoms were

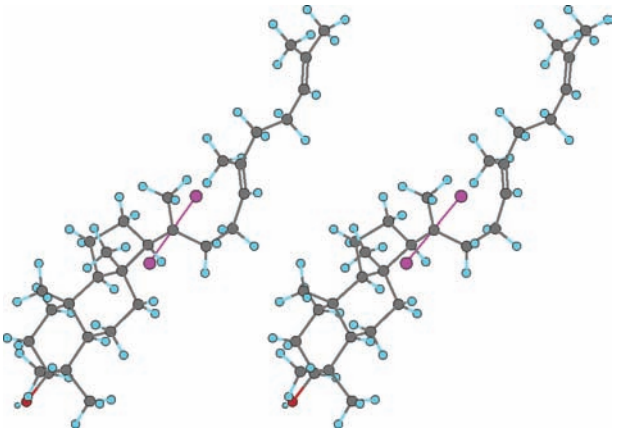


Figure 2. Stereoview of cation **7**. After geometry optimization, dummy atoms representing water (shown in magenta) were positioned 2.5 Å from the C14 cation and fixed on the axis of the empty 2p orbital. Similar stereoviews for tetracyclic and pentacyclic cations are given in the Supporting Information.

placed on the line containing C14 and C27. The proximal oxygen was extremely close to C7, C8, C26, and attached hydrogens (Figure 3A), and the rigidity of the ABC ring

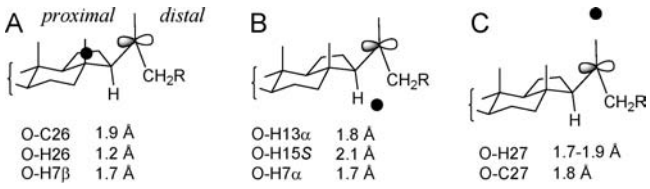


Figure 3. Steric conflicts for dummy oxygen atoms (filled circles) poised for attack on the C14 cation of **7**: O−C14, 2.5 Å (A) or 3.3 Å (B, C).

system precluded alleviating these steric conflicts by conformational adjustments. Moreover, water poised for proximal attack on **7** would sterically block the formation of rings B and C.

For water addition to vertical cation **7a** from above or below,[10] we assume that the ordered water was already replenished and fixed in position for attack when cation **7** formed. Dummy oxygen atoms 2.5 Å above and below C14 were extremely close to hydrogens on C27, C13, or C15 of **7**. The severe steric conflict could be alleviated by placing the oxygens 3.3 Å above or below C14, but the ordered water would still be close enough to the hydrogens to deprotonate cation **7** to an olefin (Figure 3B,C). Skew attack (after partial rotation about C13−C14) would also favor deprotonation. The only remaining alternative, distal attack, would not disrupt cyclization of **1** to **7**, and the side chain could fold

(7) Matsuda, S. P. T.; Wilson, W. K.; Xiong, Q. *Org. Biomol. Chem.* **2006**, *4*, 530−543.

(8) Rotation about the C13−C14 bond can occur only in the direction that places the substrate side chain in the distal position. See: Lodeiro, S.; Wilson, W. K.; Shan, H.; Matsuda, S. P. T. *Org. Lett.* **2006**, *8*, 439−442.

(9) Analogous logic for tetracycles: Xiong, Q.; Rocco, F.; Wilson, W. K.; Xu, R.; Ceruti, M.; Matsuda, S. P. T. *J. Org. Chem.* **2005**, *70*, 5362−5375.

(10) Although the absence of **5** (Figure 1) suggests that water attacks **7** before **7a** can form, our mechanistic analysis includes **7a** for completeness.

around the water (Figure S19).[11] We conclude that a cyclase active site would permit formation of little or none of the 14*S* epimer of **2** and that the C14 configuration could have been deduced a priori.

The same logic suggests that cation hydroxylation during the annulation phase of triterpene synthesis generally occurs by attack of water from the distal side. Molecular modeling of horizontal conformers of the isomalabaricadienyl, epidammarenyl, lupanyl, and hopanyl cations and **6** indicates that an ordered water positioned for proximal attack would be within 1.3 Å of multiple substrate atoms.[12] Although nondistal attack is conceivable for the dammarenyl cation[12,13] and as a very minor pathway for other intermediates, known cyclase products (notably **2** and **8**−**11**, Figure 4) that arise

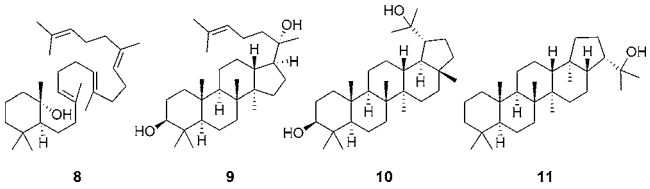


Figure 4. Selected cyclase products formed by water addition.

by water addition do so only on the distal face.[14] The strong distal bias for water addition during cyclase annulation led to configurational assignments for several triterpenes reported with ambiguous stereochemistry (see the Supporting Information).

Hydroxylation is favored over deprotonation only if the ordered water is efficiently replenished for each new substrate and if the trajectory of attack closely follows the axis of the empty 2p orbital. Thus, hydroxylation is more difficult for cyclases to carry out than cation deprotonation. This insight aligns with observations that cyclases usually terminate the cationic cascade by deprotonation[15] and that cyclases that do perform hydroxylation also generate olefinic byproducts.

Acknowledgment. The National Science Foundation (MCB-0209769), Robert A. Welch Foundation (C-1323), and Herman Frasch Foundation funded this research. Calculations were carried out in part on the Rice Terascale Cluster funded by NSF (EIA-0216467), Intel, and HP and on the Rice Cray XD1 Research Cluster funded by the NSF (CNS-0421109) in partnership with AMD and Cray. The 800 MHz NMR spectra were acquired on instrumentation purchased by the John S. Dunn, Sr. Gulf Coast Consortium for Magnetic Resonance.

Supporting Information Available: Details of experimental procedures and stereochemical assignments, NMR and GC−MS spectra, stereoviews, and analysis of constraints for water addition. This material is available free of charge via the Internet at http://pubs.acs.org.

OL070709B

(11) The ordered water that hydroxylates **7** is evidently inside the active-site cavity, where it may disrupt substrate folding to form the atypical 13*S* tricyclic cation **7** and thus prevent D-ring formation. Failure to replenish this water may result in cation **6**, the precursor of **4** and other olefins.

(12) The Supporting Information contains stereoviews and detailed analyses for these and other major structural types, including the dammarenyl cation.

(13) Only (20*S*)-**9** is common in nature, but dammar resin contains both C20 epimers: Mills, J. S.; Werner, A. E. A. *J. Chem. Soc.* **1955**, 3132−3140.

(14) **8**: (a) Hoshino, T.; Sato, T. *Chem. Commun.* **2002**, 291−301. **9**: (b) Tansakul, P.; Shibuya, M.; Kushiro, T.; Ebizuka, Y. *FEBS Lett.* **2006**, *580*, 5143−5149. The structures of **10** and **11** do not reveal the orientation of cation hydration. The orientation for **10** was deduced by isotopic labeling: (c) Kushiro, T.; Hoshino, M.; Tsutsumi, T.; Kawai, K.-i.; Shiro, M.; Shibuya, M.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 5589−5592. We determined the orientation for **11** from NMR calculations (see the Supporting Information) and isotopic labeling experiments: (d) Hoshino, T.; Nakano, S.; Kondo, T.; Sato, T.; Miyoshi, A. *Org. Biomol. Chem.* **2004**, *2*, 1456−1470.

(15) The predominance of deprotonation over hydroxylation is augmented by the nature of cyclase active sites, which contain a number of hydrophobic residues and few ordered waters: Wendt, K. U. *Angew. Chem., Int. Ed.* **2005**, *44*, 3966−3971 and references therein.