EurJOC
European Journal of Organic Chemistry

A Journal of ChemPubSoc Europe

Accepted Article

Title: Squalene-hopene cyclase: on the polycyclization reactions of squalene analogs bearing ethyl groups at positions C-6, C-10, C-15 and C-19

Authors: kazunari Takahashi, Yusuke Sasaki, and Tsutomu Hoshino

To be cited as: *Eur. J. Org. Chem.* 10.1002/ejoc.201800010

Link to VoR: http://dx.doi.org/10.1002/ejoc.201800010

Supported by ACES Asian Chemical Editorial Society

WILEY-VCH

10.1002/ejoc.201800010

Squalene-hopene cyclase: on the polycyclization reactions of squalene analogs bearing ethyl groups at positions C-6, C-10, C-15 and C-19

Kazunari Takahashi,[+] Yusuke Sasaki[+] and Tsutomu Hoshino*

Graduate School of Science and Technology and Department of Applied Biological Chemistry, Faculty of Agriculture, Niigata University, Ikarashi 2-8050, Nishi-ku, Niigata 950-2181, Japan

*Corresponding author. Prof. Dr.
E-mail address: hoshitsu@agr.niigata-u.ac.jp
http://www.agr.niigata-u.ac.jp/~satot/index.html

[+] These authors contributed equally to this work.

Supporting Information for this article is available under WWW. https……

 10.1002/ejoc.201800010

Abstract

Squalene-hopene cyclase (SHC) converts acyclic squalene into the 6,6,6,6,5-fused pentacyclic triterpene hopene and hopanol. Enzymatic reactions of squalene analogs bearing ethyl groups in lieu of methyl groups at positions C-6, C-10, C-15 and C-19 were examined in order to investigate whether the larger ethyl-substituents (C_1-unit increment) are accepted as substrates and to investigate how these substitutions affect polycyclization cascades. Analog 6-ethylsqualene **19a** was not cyclized, indicating that substitution with the bulky group at C-6 completely inhibited the polycyclization reaction. In contrast, 19-ethylsqualene **19b** afforded a wide spectrum of cyclization products, including mono-, bi-, tetra-, and pentacyclic products in a ratio of 6:6.3:1:2. Production of tetra- and pentacyclic scaffolds suggests that the reaction cavity for the D-ring formation site is somewhat loosely packed and can accept the 19-ethyl group, and a robust hydrophobic interaction exists between the 19-ethyl group and the binding site. In contrast to **19b**, 10-ethylsqualene **20a** and 15-ethylsqualene **20b** afforded mainly mono- and bicyclic products, that is, the polycyclization cascade terminated prematurely at the bicyclic reaction stage. Therefore, the catalytic domain for the 10- and 15-methyl binding sites are tightly packed and cannot fully accommodate the Et substituents. The cyclization pathways of ethyl-substituted substrates are compared between SHC and lanosterol and β-amyrin synthases.

Introduction

The enzymatic cyclization of acyclic squalene **1** and 2,3-oxidosqualene to form polycyclic triterpenes have fascinated chemists and biochemists for over half a century.[1, 2] The X-ray crystallographic structure of the squalene-hopene cyclase from *Alicyclobacillus acidocaldarius* (*Aa*SHC) was reported in 1997.[3] *Aa*SHC catalyzes the conversion of **1** into the pentacyclic triterpenes hopene **2** and hopanol **3** at a ratio of *ca.* 5:1 (Scheme 1).[2a, 2b] This polycyclization cascade is attained by a single enzyme. Compound **1** is folded in all pre-chair conformations inside the reaction cavity and cyclized in a regio- and stereospecific fashion via a series of carbocationic intermediates, leading to the formation of five new C-C bonds and nine chiral centers. Studies on *Aa*SHC have revealed that the polycyclization reaction consists of 8 reaction steps (Scheme 1): [2b] (1) first, cyclization to form A-ring **4** by proton attack on the terminal double bond, donated by the DXDD motif;[4] (2) second, ring closure to yield the B-ring (6/6-fused A/B ring system **5**);[5,6] (3) third, cyclization to yield the 5-membered C-ring (6/6/5-fused A/B/C-tricyclic ring system **6**) by Markovnikov closure,[6, 7] which then undergoes (4) ring expansion to form the 6-membered C-ring

 10.1002/ejoc.201800010

(6/6/6-fused tricyclic ring system **7**);[7] (5) fifth, cyclization to yield the thermodynamically favored 5-membered D-ring (6/6/6/5-fused A/B/C/D ring system **8**, 17-*epi*-dammarenyl cation) [8-10] followed by (6) the second ring enlargement process to form the 6-membered D-ring (6/6/6/6-fused A/B/C/D-ring system, prohopanyl cation **9**) [8-12] and (7) the last ring closure process to construct the 6/6/6/6/5-fused A/B/C/D/E-ring system (**10**, hopanyl cation),[12] and (8) the final deprotonation reaction of **10** to introduce the double bond.[13] Recently, the Hauer group has succeeded in creating the *Aa*SHCs with novel catalytic activities by site-directed mutagenesis, which have enabled the cyclization reactions of truncated squalene and terpene-like analogs to yield unnatural products.[2i]

<Scheme 1>

We have reported the polycyclization reactions of various norsqualene analogs by the native *Aa*SHC; in these analogs, the methyl positions of **1** are substituted with hydrogen atoms. Squalene **1** is a symmetrical molecule and possesses two isopropylidene moieties at both terminal positions. We have revealed that the isopropylidene moiety is essential for initiating polycyclization by *Aa*SHC. [13] In previous paper, we reported the cyclization products of 26-norsqualene **11**. [14] This compound possesses isopropylidene moieties at both left and right termini, and thus, two cyclization pathways are possible, i.e., from the left (path *a*) and right side (path *b*). Therefore, two chemical names were proposed for this compound based on the cyclization pathway: 26-norsqualene **11a** for path *a* and 29-norsqualene **11b** for path *b*, as shown in Scheme 2A. Substrate **11a** afforded **12**, and **11b** afforded **13** and **14**. [14] The structures of all the products indicate that both **11a** and **11b** are folded in a normal folding conformation, as shown in Scheme 1. Thus, the substitutions of the less-bulky hydrogen atom at C-6 and C-19 had little influence on the polycyclization cascade. On the other hand, 27-norsqualene **12a** generated the unprecedented triterpene skeletons **15** and **16**, which consist of 6/5+5/5 and 6/5+5/5+6 ring systems, respectively. [15] The structures of **15** and **16** indicate that **12a** underwent the unusual folding conformation, as depicted in Scheme 2. In contrast, 28-norsqualene **12b** afforded the pentacyclic hopane skeletons **17** and **18** as the enzymatic products. Product **17** was produced from the all-chair conformation, as shown in Scheme 1, but product **18** was generated from a chair/chair/chair/chair/boat folding conformation.[15] The results from **12a** and **12b** indicated that the decrease in steric size at C-10 of **1** had a greater effect on the folding conformation than that at C-15.

 10.1002/ejoc.201800010

<Scheme 2>

However, there is no report regarding the cyclization reactions of squalene analogs **19** and **20**, which bear larger substituents (ethyl moieties, C_1-increment) at the C-6 or C-19 and the C-10 or C-15 positions. Analog **19** has an Et group at C-6 or C-19 (**19a** or **19b**, respectively) based on the cyclization pathways *a* and *b*, respectively. Homolog **20** possesses Et groups at C-10 or C-15 (**20a** or **20b**, respectively). Analogs **19** and **20** were synthesized and incubated with *Aa*SHC. Introduction of the Et moiety at C-6 (**19a**) gave no enzymatic product, while that at C-19 (**19b**) afforded mono-, bi-, tetra- and pentacyclic scaffolds. Replacement with an Et group at C-10 (**20a**) and C-15 (**20b**) led to the production of only mono- and bicyclic compounds with the exception of the generation of a tricyclic product from **20a**. Herein, we report the detailed experimental results and discuss the mechanism of the cyclization of **19** and **20**, which exhibit larger steric volumes (C_1-increment). This investigation provides important information on the molecular recognition mechanism between *Aa*SHC and squalene substrate, and the differences in the polycyclization results between lanosterol and β-amyrin synthases are discussed.

Results

Syntheses of ethyl-substituted analogs 19 and 20

Squalene **1** was treated with *m*CPBA (*m*-chloroperbenzoic acid) in CH_2Cl_2 to produce an epoxide mixture, which was then treated with H_5IO_6, yielding a mixture of C_{17} **21**, C_{22} **22**, and C_{27} aldehydes.[13, 16] Scheme 3 shows the synthetic designs of **19** and **20**. Synthetic intermediates of **23** were prepared from aldehyde **21** according to the published protocol.[17] Aldehyde **21** was subjected to a Wittig-Horner reaction using triethyl 2-phosphonobutyrate to yield an *E/Z*-mixture of alkenes, which were separated by SiO_2 column chromatography to isolate the *E*-isomer. The ethyl ester thus obtained was reduced with DIBAL-H to yield the corresponding alcohol, which was then converted to bromide **23** using PBr_3. Geraniol **24** was treated with PBr_3 to yield geranyl bromide. The solution of bromide in DMF was slowly added to sodium benzenesulfinate in DMF, yielding geranyl phenylsulfone **25**. Compounds **23** and **25** were subjected to a coupling reaction using *n*-BuLi, yielding the phenylsulfone derivative of the ethyl-substituted analog **26**. To remove the phenylsulfonyl group, **26** was treated with Super-Hydride reagent to synthesize the required C_{31} analog **20**. To synthesize analog **19**, aldehyde **22** was subjected to the same reactions as the preparation of **23** from **21**, affording bromide **26**. Commercially available 3-methylbut-

 10.1002/ejoc.201800010

2-en-1-ol was transformed to the bromide and then to the corresponding phenylsulfone derivative **27** by using the same reactions as the preparation of **25** from **24**. The coupling reaction of **26** with **27** and the subsequent removal of the phenylsulfone group were carried out, yielding the C_{31} analog **19**. The synthetic experiments and the NMR data of the synthetic intermediates and Et-substituted analogs **19** and **20** are described in the Supporting Information (Fig. S1).

<Scheme 3>

Enzymatic reactions of 19 and 20 with native *Aa*SHC

A buffered reaction mixture (2.0 mL, pH 6.0, 60 mM citrate buffer) consisting of Triton X-100 (0.2%, *w*/*v*), **1** [200 μg, **19** or **20**], and purified histidine-tagged pET26b *Aa*SHC (5 μg)[18] was incubated at 55 °C for 6 h.[18,19] Then, 15% KOH/MeOH (7.5 mL) was added, and the resulting mixture was heated to 70–80 °C for 30 min. The lipophilic materials were extracted with hexane, and geranylgeraniol (5 μg) was added to the hexane extract as the internal standard. Triton X-100 present in the extract was removed by elution through an SiO_2 column (hexane/EtOAc, 100:10), and the eluent was evaporated to dryness. Then, hexane (150 μL) was added to the residue, and the resultant solution (1.0 μL) was subjected to GC analysis (GC: gas chromatography) to quantify the lipophilic materials.

Fig. 1 shows the gas chromatograms of the hexane extracts obtained by incubating **1** (A), **19** (B) and **20** (C) with the purified His-tagged *Aa*SHC. In Fig. 1A, no peak corresponding to substrate **1** was observed (A), indicating that **1** was completely converted to **2** and **3** under the incubation conditions described above. Incubation of **19** afforded a broad product spectrum, as shown by the products numbered **28**–**33**; however, a significantly large peak corresponding to substrate **19** was observed (B), indicating that the conversion of **19** was markedly decreased relative to that of **1**. Fig. 1C shows that substrate **20** also exhibited significantly low conversion (a large peak corresponding to substrate **20**) despite many products **34**–**42** having been generated. Consequently, substitution of the bulkier moiety (increment of C_1-unit) at positions C-6, C-10, C-15 and C-19 resulted in significant perturbation of the polycyclization reaction.

<Fig. 1>

Isolation and structures of enzymatic products 28–33 from substrate 19

To isolate products and determine their structures, we cultured *Escherichia coli* BL21(DE3) harboring pET3a-*Aa*SHC and used the cell-free extracts for large scale incubation experiments.[19] Substrate analog **19** (27 mg) was incubated with the cell-free extract (60 mL), which was prepared from a 1.2-L culture, at optimal conditions (60 °C,

 10.1002/ejoc.201800010

pH 6.0) for 16 h. [19] After saponification with a solution of 15% KOH/MeOH, the lipophilic materials were extracted with hexane, and then, the Triton X-100 was removed from the hexane extract by a short SiO_2 column chromatographic purification (hexane:EtOAc=100:20). Careful SiO_2 column chromatography (activated by heating SiO_2 at 170 °C for 3 h), eluting with hexane, afforded **28**, **30** and **32** in pure states. A fraction containing **29** and impurities was subjected to normal phase HPLC with 100% hexane to afford pure **29**. Fractions enriched with **31** and **33** were subjected to argentation SiO_2 column chromatography (5% $AgNO_3$, 100% hexane), which afforded pure **31** and **33**. Final purifications of all products were performed by normal phase HPLC (100% hexane).

Substrates **19** and **20** possess seven methyl groups and one ethyl group on the squalene backbone. In the ^{1}H-NMR spectrum of product **28** (400 MHz, $CDCl_3$), four vinylic Me groups were found at δ_H (ppm) 1.59 (3H, s), 1.61 (6H, s, Me x 2) and 1.68 (3H, s), and one olefinic Et group was observed at δ_H 0.959 (3H, t, *J*=7.6 Hz, Me-31) and 2.03 (2H, m, H-28), suggesting that **28** is a monocyclic product and that the Et group is present on the double bond of **19**. Two aliphatic Me groups (δ_H 0.911, 3H, s for Me-23; 0.833, 3H, s for Me-24) and one methylidene group (δ_H 4.54, 1H, s and δ_H 4.75, 1H, s for H-25) were also detected. Distinct HMBC cross peaks were observed for Me-23/Me-24/C-5 (δ_C 53.6, d)/C-3 (δ_C 36.4, t)/C-4 (34.9, s) and for H-25/C-5/ C-6 (δ_C 149.4, s). Detailed analyses of 2D NMRs, including ^{1}H-^{1}H COSY, HOHAHA, NOESY, HSQC and HMBC spectra (Figs. S2.2─S2.8, Supporting Information), in addition to DEPTs data demonstrated that the structure of **28** is as depicted in Fig. 2. Complete assignments of ^{1}H- and ^{13}C-NMR spectra are shown in Fig. S2.9.

In the ^{1}H-NMR spectrum of product **29** (600 MHz, C_6D_6), five olefinic Me groups and one olefinic Et group were detected in addition to five olefinic protons (δ_H 5.38, t, *J*=6.8 Hz, 2H, H-17 & H-21; δ_H 5.47, 2H, s, H-10 & H-13; 5.59, 1H, s, H-3). Since seven olefinic Me groups and one Et group, in addition to six olefinic protons, are involved in substrate **19**, the cyclization product of **19** is likely **29**. Indeed, two aliphatic Me groups were found at δ_H 0.979 (3H, d, *J*=6.8 Hz, Me-25) and 1.01 (3H, s, Me-24), suggesting that **29** is a monocyclic product with one double bond in the ring. Olefinic Me-23 (δ_H 1.79, 3H, s) had clear HMBC cross peaks for C-3 (δ_C 124.5, d), C-4 (δ_C 139.7, s) and C-5 (δ_C 40.7, s), and Me-24 showed distinct HMBC cross peaks for C-4, C-5 and C-6 (δ_C 33.5, d). Detailed HMBC analyses (Fig. S3.9) showed that **29** possesses a 1,5,6-trimethylcyclohexene moiety (IUPAC numbering, Fig. S3.10B). This ring system was further confirmed by electron impact mass spectrometry (EIMS), with fragment ion *m*/*z* 123 as a base peak (Fig. S3.1). The relative stereochemistry of the

 10.1002/ejoc.201800010

cyclohexane ring was determined by the NOESY spectrum (Fig. S3.6). Distinct NOEs of Me-24/H-6 and Me-25/H-7 were observed, but no NOE was observed for H-6/H-7 or for Me-25/Me-24 (see Figs. S3.9 and S3.10A). These findings strongly indicate that the relative configurations of positions 5 and 6 of the cyclohexene ring all possess *R* stereochemistry, indicating that substrate **19** was folded in a boat conformation but not in a chair conformation (see Fig. S3.10A and Scheme 4E). Thus, the structure of product **29** was determined as shown in Fig. 2 and in Figs. S3.9 and S3.10. We have reported that this 1,5*R*,6*R*-trimethylcyclohexene ring (IUPAC numbering, Fig. S3.10B) was produced by incubating **1** with the L607F SHC variant. We have proposed the name neoachillapentaene for the structure of the fundamental skeleton of **29**.[19c] This unusual product with the same relative stereochemistry as **29** was also obtained from the incubation experiment of 10-ethyl-2,3-oxidosqualene with hog-liver lanosterol cyclase (see **44** shown in Fig. S14.1).[20] The absolute stereochemistry at positions 5 and 6 should be verified by other methods different from NMR.

The ^{1}H-NMR spectrum of product **30** (400 MHz, $CDCl_3$) showed that three olefinic Me groups were identified: δ_H 1.57 (3H, s, Me-27), 1.60 (3H, s, Me-30) and1.68 (3H, s, Me-29). In addition, one olefinic Et group was observed: δ_H 0.963 (3H, t, *J*=7.6 Hz, Me-31) and 2.03 (2H, m, H-28). Three aliphatic methyl groups were also identified: δ_H 0.666 (3H, s, Me-25), δ_H 0.800 (3H, s, Me-24) and δ_H 0.868 (3H, s, Me-23). In addition, one methylidene group was observed: δ_H 4.54 (1H, s, H-26) and 4.82 (1H, s, H-26). These findings suggest that **30** is a bicyclic compound, which was further confirmed by the following HMBC correlations (Fig. S4.9): H-26/ C-8 (δ_C 148.8, s) and H-26/C-9 (δ_C 56.2, d), Me-25/C-9, Me-23/C-5 (δ_C 55.6, d), Me-24/C-5, Me-25/C-5. A strong NOE was observed between H-5 (δ_H 1.08, 1H, dd, *J*=12.8, 2.8 Hz) and H-9 (δ_H 1.61, 1H, m). Complete NMR assignments are shown in Fig. S4.9. The structure of **30** thus determined is shown in Fig. 2 and Fig. S4.9.

Product **31** showed two olefinic Me groups (600 MHz, C_6D_6): δ_H 1.84 (3H, s, Me-26) and 1.74 (3H, s, Me-27). An Et group was connected to C-20 (δ_C 39.8, d), which was determined by the HMBC correlation between Me-31 (δ_H 0.959, 3H, t, *J*=7.1 Hz) and C-20. The C-20 signal appeared at significantly higher field than that (δ_C 140.8, s) of the corresponding position of **19**. Thus, the Et group is not present on the double bond of the squalene backbone, indicating that olefinic Et changed to aliphatic Et. The presence of the Et group on C-20 was further confirmed by the following HOHAHA connectivity: Me-31/H-21 (δ_H 1.46, 1H, m; 1.51, 1H, m)/H-20 (δ_H 2.42, 1H, m). Thus, product **31** is a tetracyclic compound. Strong HMBC cross peaks were observed between H-21 and C-17 (δ_C 132.7, s), and between Me-30 (δ_H 1.31, 3H, s) and C-13 (δ_C

 10.1002/ejoc.201800010

142.0, s), indicating that the double bond is situated at C-13 and C-17. Further analysis of the HMBC spectrum (Figs. S5.8 and S.5.9) allowed us to propose the fundamental structure of **31** as dammara-13(17), 24-diene. Unambiguous NOEs were observed for Me-28 (δ_H 1.04, 3H, s)/H-5 (δ_H 0.940, 1H, m), H-5/H-9 (δ_H 1.60, 1H, m) and H-9/Me-30 (δ_H 1.31, 3H, s), and thus, the complete structure, including stereochemistry, can be proposed as shown in Fig. 2 and Fig. S5.9, except for the C-20 stereochemistry.

The ^{1}H-NMR spectrum (400 MHz, $CDCl_3$) of product **32** showed four olefinic Me groups: δ_H 1.60 (3H, s, Me-30), 1.61 (3H, s, Me-27), 1.68 (3H, s, Me-29), 1.71 (3H, s, Me-26). In addition, the spectrum showed one olefinic Et group: δ_H 0.980 (3H, t, J=6.8 Hz, Me-31) and 2.03 (2H, m, H-28). Furthermore, three aliphatic Me groups were found at δ_H 0.852 (3H, s, Me-23), 0.874 (3H, s, Me-24) and 0.741(3H, s, Me-25). These data suggest that **32** may be monocyclic. However, the HMBC cross peaks were clearly observed for Me-25/C-9 (δ_C 54.2, d), Me-26/C-9, Me-26/C-7 (δ_C 122.0, d) and Me-26/C-8 (δ_C 135.6, s), indicating that **32** was indeed a bicyclic compound and that double bond is situated at positions C-7 and C-8. Detailed analysis of 2D NMR spectra revealed that the complete structure of **32** is as shown in Fig. 2 and Fig. S6.9. Structural difference between **30** and **32** is found only at the double bond position.

Product **33** possessed only one double bond, which was observed at δ_H 5.11 (1H, s, H-29) and δ_H 5.07 (1H, s, H-29) in the ^{1}H-NMR (600 MHz, C_6D_6, Fig. S7.2) spectrum and δ_C 109.9 (t, C-29) and δ_C 148.9 (s, C-22) in the ^{13}C-NMR (150 MHz, C_6D_6, Fig. S7.3) spectrum. The two olefinic protons (H-29) were correlated with δ_C 109.9 (t) in the HSQC spectrum (Fig. S7.7) and had HMBC cross peaks for C-22 (Fig. S7.8). Thus, the double bond is assigned to a methylidene group. No other double bond was observed. This finding definitively demonstrated that product **33** was produced by the complete cyclization reaction, suggesting that **33** has a hopene scaffold. The detailed 2D NMR analyses are shown in Fig. S7.9. In the HMBC spectrum, H-21 (δ_H 2.72, 1H, m) had clear cross peaks with the following carbons: C-22, C-29, C-30 (δ_C 25.7, q), C-17 (δ_C 56.3, d), and C-20 (δ_C 28.1, t). H-21 exhibited unambiguous HOHAHA correlations for the following protons: H-17/H-16/H-15/H-20/H-19 (Fig. S7.5). The HMBC and HOHAHA data strongly indicated that **33** possesses a hopene skeleton with a 5-membered E-ring. An Et group is attached to the C-18 position, as the Me protons (δ_H 1.02, 3H, t, J=7.4 Hz, Me-31) displayed a strong HMBC cross peak for C-18 (δ_C 47.4, s). Taking NOE data into consideration, the whole structure of product **33,** including the stereochemistry, can be proposed as shown in Fig. 2 and Fig. S7.9.

<Fig. 2>

 10.1002/ejoc.201800010

Isolation and structures of enzymatic products 34─42 from substrate 20

Substrate analog **20** (40 mg) was incubated with the cell-free extract (100 mL) at optimal conditions (60°C, pH 6.0) for 16 h. The extraction of the lipophilic materials from the incubation mixture was carried out with the same method that was used for the incubation of **19**. By careful SiO_2 column chromatography with a step-wise elution (hexane/EtOAc, 100:0–100:5), four partially purified fractions were obtained. Each fraction was further purified by 5% or 10% $AgNO_3$-SiO_2 column chromatography (hexane/EtOAc, 100: 0─100:3). Final purification was conducted by normal phase HPLC (100% hexane). Complete separation of **34** and **35**, **36** and **38** and **39** and **40** was unsuccessful. Therefore, each mixture of two compounds was directly subjected to NMR analyses without further purification.

Fig. S8.1 shows the GC-EIMS results of a mixture of **34** and **35**, which exhibited identical fragment patterns, suggesting that the two products have similar structures and that the ratio of the yields of **34** to **35** was ca. 1 to 1.5. The ^{1}H-NMR spectrum (600 MHz, C_6D_6) is shown in Fig. S8.2. Two aliphatic Me groups were found with high intensity peaks: δ_H 0.974 (3H, s, Me-24) and 1.08 (3H, s, Me-23). In addition, an Et group was identified (δ_H 1.11, 3H, t, J=5.2 Hz, Me-31;δ_H 2.20, 2H, m, H-26). These intense proton signals are accompanied by similar proton signal patterns but with lower intensities by ca. 1/1.5; the lower signals were as follows: δ_H 0.983 (3H, s, Me-24), 1.08 (3H, s, Me-23), and δ_H 1.14 (3H, t, J=5.2 Hz, Me-31), and 2.23 (2H, m, H-27) for the Et group (see Fig. S8.2B). Thus, we were able to differentiate the ^{1}H-NMR signals of **35** (higher signals) from those of **34** (lower signals). Methylidene protons (δ_H 4.99, s; 4.82, s: H-25), which were correlated with C-25 (δ_C 109.3, t) in the HSQC spectrum and were not distinguishable between **35** and **34**, showed clear cross peaks with C-6 (δ_C 149.5, s, for both **35** and **34**), C-1 (δ_C 32.8, t, for both **35** and **34**) and C-5 (δ_C 53.97, d, for **35**; δ_C 54.16, d, for **34**). The ^{13}C-NMR spectrum (expanded) is shown in Figs. S8.3B and S8.3C. Both Me-23 and Me-24 of **35** and **34** had strong HMBC correlation with C-3 (δ_C 36.6, t, for both **35** and **34**). The COSY and HOHAHA spectra revealed the ^{1}H-^{1}H spin couplings of H-1/H-2/H-3, as shown in Fig. S8.9. Further, the ^{1}H-^{1}H spin coupling networks H-5/H-7/H-8 were found from the HOHAHA spectrum. Observation of the strong HMBC correlation between Me-26 (δ_H 1.77, s) and C-8 (δ_C 35.6, t), which was found only for **34**, indicated that **34** has a Me group at C-9 (δ_C 135.8, s). Thus, **34** and **35** are monocyclic compounds, and the complete structures of **35** and **34** are as depicted in Fig. 2. The detailed assignments are shown in Fig. S8.9. As described below, the Et group of product **35** is positioned at C-9 (δ_C 141.1, s), based on the cyclization pathway (Scheme 4).

 10.1002/ejoc.201800010

The GC-EIMS spectrum of the mixture of products **36** and **38** is shown in Fig. S9.1. The two products showed similar fragment patterns, which is suggestive of the similar structures of these two products. The yield of **38** was significantly higher than that of **36**. The ratio was estimated to be ca. 4:1 by integration of the Me-25 signals in the ^{1}H-NMR spectrum (600 MHz, $CDCl_3$, see Fig. S9.2A and B), and thus, we could easily differentiate the signals of **38** from those of **36**. Three aliphatic Me protons and four vinylic Me protons were found in the ^{1}H-NMR spectrum of **38** (see Fig. S9.2B): δ_H 0.615 (3H, s, Me-25), 0.786 (3H, s, Me-24), 0.867 (3H, s, Me-23); δ_H 1.56 (3H, s, Me-27), 1.60 (6H, s, Me-28 and Me-30), 1.68 (3H, s, Me-29), but the Et signal of substrate **20** was not observed. Instead, a doublet Me proton signal (δ_H 1.65, d, J=6.6 Hz, Me-31), which was correlated with the olefinic proton (δ_H 5.04, 1H, q, H-26) in the COSY spectrum, appeared, definitively indicating the presence of a CH_3CH= group in **38**. The following ^{1}H-^{1}H spin coupling networks were revealed by analyzing the COSY and HOHAHA spectra: H-5 (δ_H 1.08, dd, J=8.4, 2.0 Hz)/H-6 (δ_H 1.20, m; 1.70, m)/H-7 (δ_H 1.56, m; 2.78, m). Olefinic H-26 had distinct HMBC cross peaks for C-7 (δ_C 30.4, t) and C-9 (δ_C 57.0, d). Me-25 also showed a strong HMBC contour peak for C-9. The expanded spectrum of the ^{13}C-NMR is shown in Fig. S9. 3B. Thus, **38** is a bicyclic compound as shown in Fig. 2 and Fig. S9.9. In contrast, **36** showed the presence of methylidene protons (δ_H 4.82, s; 4.53, s, CH_2-26) with ca. one-fourth of the intensity of H-26 of **38**. Furthermore, **36** had an olefinic Et group (δ_H 0.948, t, J=7.8 Hz, Me-31 and δ_H 1.99, m, CH_2-27). Thus, the structure of **36** could be proposed as shown in Fig. 2 and Fig. S9.9. The detailed assignments are depicted in Fig. S9.9.

The ^{1}H-NMR spectrum (600 MHz, $CDCl_3$, see Fig. S10.2) showed that the isolated **37** was a single product (not a mixture of two products, as observed for **34**+**35** and for **36**+**38**). Three olefinic Me groups were found: δ_H 1.68 (3H, s, Me-29), 1.604 (3H, s, Me-30) and 1.559 (3H, s, Me-28). Three aliphatic Me groups were observed: δ_H 0.905 (3H, s, Me-24), 0.869 (3H, s, Me-23) and 0.734 (3H, s, Me-25). In the HMBC spectrum (Fig. S10.8), the following correlations were observed: Me-23/Me-24/C-3 (δ_C 42.8, t)/C-4 (δ_C 32.9, s)/C-5 (δ_C 50.7, d) and Me-25/C-1(δ_C 40.4, t)/C-5/C-9 (δ_C 58.8, d)/C-10 (δ_C 34.9, s). Olefinic proton H-7 (δ_H 5.20, 1H, brs) had a distinct HMBC contour with C-9, and further, this proton clearly showed the following COSY and HOHAHA correlations: H-7/H-6 (δ_H 1.88, 1H, m; 2.08, m)/H-5 (δ_H 1.22, 1H, m). The H-9 (δ_H 1.95, m) also had HOHAHA cross peaks for H-11 and H-12. These data suggest that the double bond is situated at C-7 and C-8. The proton signals of an Et moiety were found: δ_H 0.772 (3H, t, J=7.4 Hz) for Me-31; δ_H 1.35 (1H, m) and δ_H 1.52 (1H, m) for H-26. However, Me-31 had strong HMBC correlations with C-13 (δ_C 50.7, s) and C-26 (δ_C

 10.1002/ejoc.201800010

30.4, t). The chemical shift of C-13 definitively demonstrates that the Et group of **37** is not placed on the double bond that is seen in substrate **20**. The doublet Me group (δ_H 0.823, 3H, d, J=6.7 Hz, Me-27) showed a clear HMBC cross peak for C-13. Consequently, product **37** is a 6,6,5-fused tricyclic compound with a double bond at C-7 and C-8; the double-bond position was further established by definitive HMBC contour of H-26 with C-8 (δ_C 147.1, s). Distinct NOEs were found for H-5/H-9 and H-7/H-14 (δ_H 1.38, 1H, m), indicating that the Et group is β-oriented, and the side chain is α-configured. Therefore, the complete structure of **37** is as depicted in Fig. 2 and Fig. S10.9.

Products **39** and **40** were not distinguishable by the GC analysis (Fig. 1), and thus, the yield ratio of **39** and **40** was not determined from the GC chromatogram. The ^{1}H-NMR spectrum of a mixture of **39** and **40** (600 MHz, $CDCl_3$) showed that the yield ratio was ca. 1:1 (Fig. 11.2B). Two Et groups were found in a 1:1 ratio; the proton signals of δ_H 0.996 (3H, t, J=7.8 Hz, Me-31) and δ_H 2.02 (2H, m, H-27) were observed for the Et group of **39**; δ_H 1.01 (3H, t, J=7.2 Hz, Me-31) and 2.03 (2H, m, H-26) were assigned to the Et group of **40**. A pair of three aliphatic Me groups were also observed in a 1:1 ratio: δ_H 0.871 (s, Me-24), 0.849 (s, Me-23) and 0.740 (s, Me-25) for **39**; δ_H 0.878 (s), 0.857 (s, Me-23) and 0.741 (s, Me-25) for **40**, but these assignments may be exchangeable between **39** and **40**. The signals of Me-23, Me-24 and Me-25 were assigned by HSQC and HMBC analyses (Figs. S11.7 and S11.8), as depicted in Fig. S11.9. The HMBC cross peaks from Me-23, Me-24 and Me-25 also enabled the unambiguous assignment of C-1, C-3, C-5 and C-9. The COSY and HOHAHA analyses (Fig. S11.4 and S11.5) revealed the presence of the following cross peaks: H-5/H-6/H-7 (δ_H 5.38, brs); H-7 is an olefinic proton. Definitive NOEs of H-7/Me-26 (δ_H 1.71, s) for **39** and H-7/Me-31 (δ_H 1.01, t, J=7.2 Hz) for **40** (Fig. S11.6) gave important information for differentiating **39** from **40**, that is, Me-26 of **39** and the Et group of **40** are positioned at C-8, which was further supported by the distinct HMBC correlations of Me-26/C-7 (δ_C 122.1, d)/C-8 (δ_C 135.6, s)/C-9 (δ_C 54.4, d) for **39** and CH_2-26 (δ_H 2.03, m, 2H)/C-7(δ_C 119.7, d)/C-8 (δ_C 141.2, s) for **40**. Me-25 of **39** and **40** had strong HMBC correlations with C-9 of **39** and **40**. Consequently, **39** and **40** were determined to be bicyclic products as shown in Fig. 2. Structural differences between **39** and **36** and between **40** and **38** were found only at the double-bond positions.

Fortunately, product **41** was successfully isolated in a pure state, albeit at a low yield (Fig. 1). SiO_2 TLC chromatography showed that **41** was a highly polar compound; the R_f value was ca. 0.4 with the mixed solvent (hexane/EtOAc,100:10), but no movement was observed with hexane (100%), while products **34** and **35** showed R_f

 10.1002/ejoc.201800010

values of ca. 0.6 with hexane (100%). The ^{1}H-NMR spectrum of **41** (600 MHz, C_6D_6, see Fig. S12.2) showed the presence of three aliphatic Me and four olefinic Me groups: δ_H 0.872 (3H, s, Me-24), 1.06 (3H, s, Me-23), 1.20 (3H, s, Me-25), 1.69 (3H, s, Me-30), 1.74 (3H, s, Me-28), 1.81 (3H, s, Me-29) and 1.85 (3H, s, Me-26). In addition, the presence of an Et group was observed: δ_H 1.12 (3H, t, *J*=7.4 Hz, Me-31) and 2.19 (2H, q, *J*=7.4 Hz, H-27), indicating that **41** is a monocyclic compound and that the Et group is situated on the double bond of the squalene backbone. The ^{13}C-NMR spectrum (150 MHz, C_6D_6, see Fig. S12.3) indicated that a quaternary alcoholic carbon (δ_C 73.5, s, C-6) is present in **41**. Me-25 showed distinct HMBC cross peaks for C-1 (δ_C 43.9, t), C-5 (δ_C 56.9, d) and C-6. Me-23 and Me-24 had strong HMBC correlations with C-3 (δ_C 41.8, t), C-4 (δ_C 35.6, s) and C-5. The H-5 (δ_H 1.22, 1H, m) was assigned by the HSQC cross peak from C-5. In the COSY and HOHAHA spectra (Fig. S12.4 and S12.5B), the following ^{1}H-^{1}H connectivity was found: H-5/H-7 (δ_H 1.56, 1H, m; 1.85, 1H, m)/H-8 (δ_H 2.33, 1H, m; 2.46, 1H, m). Strong HMBC cross peaks for Me-26/C-8 (δ_C 43.3, t) and Me-26/C-9 (δ_C 136.5, s) provided important information that the substituent at C-9 is an Me group and not an Et group, as depicted in Fig. 2 and Fig. S12.9. The detailed analyses, including 2D NMR analyses, are shown in Fig. S12.9.

Production yield of **42** was also trivial (Fig. 1). Product **42** was a highly polar compound (same R_f value as **41** by SiO_2 TLC), indicating that **42,** like **41,** has a hydroxyl group. In spite of careful purification, the isolation of **42** was, unfortunately, unsuccessful; sample loss occurred during several purification steps, leading to a low yield. However, the EIMS spectrum of **42** was identical to that of **41**, including the fragment patterns (Fig. S13), which suggests that the structure of **42** is as shown in Fig. 2. This idea is further supported by the finding that the EIMS of **34** was identical to that of **35**, although the structure of **34** is different from that of **35** with respect to the Et-substituent position. Moreover, taking the cyclization mechanism into consideration, **42** has Et and Me groups at C-9 and C-14, respectively, while the structure of **41** was established by the NMR analyses to have Et and Me groups at C-14 and C-9, respectively, as described above (see Fig. 2).

Discussion

Cyclization pathways of 19a, 19b, 20a and 20b by *Aa*SHC

As shown in Fig. 2, no cyclization products from **19a** was observed, or if they were present they were below the limit of detection, thus indicating that the introduction of a larger Et group at C-6 prevented the polycyclization reaction. In contrast, the analog

 10.1002/ejoc.201800010

that was Et-substituted at C-19 (**19b**) underwent cyclization reactions to yield a wide spectrum of mono-, bi-, tetra- and pentacyclic products. The cyclization mechanisms are depicted in Scheme 4. The cyclization reaction of **19b** halted at the monocyclic cation (**4'**), the structure being analogous to cation **4**. The proton elimination from Me-25 of **4'** (path *c*) afforded product **28**. The same deprotonation reactions of cation **4'** (path *c*) from **20a** and **20b** gave products **35** and **34**, respectively. Attack by a water molecule on cation **4'** yielded **41** and **42** (path *d*). Further cyclization of **4'** furnished the bicyclic cation **5'** (path *e*), which was structurally similar to cation **5**, which underwent deprotonation at Me-26 (path *f*), yielding products **30** and **36** from **19b** and **20b**, respectively. Deprotonation of H-7 of cation **5'** afforded **32** and **39** from **19b** and **20b**, respectively. Analog **20a,** with the bulky Et group at C-10, also underwent further cyclization to give the bicyclic cation **5''**. Deprotonation of the Et group (path *f*) and that of H-7 (path *g*) generated products **38** and **40**, respectively. Further cyclization of **5''** afforded the 6,6,5-fused tricyclic cation **6''** (path *h*). The successive 1,2-shift reactions occurred as follows (path *i*): H-13→cation C-14, Et at C-8 → C-13, and deprotonation of H-7α. The Et and H-7α are arranged in anti-parallel fashion, thus the deprotonation reaction occurs from H-7α, but not from H-7β. Notably, the tricyclic product **37** was generated only from **20a** and not from **20b**, that is, **20b** yielded only mono- and bicyclic products. Therefore, the bulky Et substituent at C-10 is, to an extent, better tolerated than that at C-15; the hydrophobic interaction between the alkyl group and the binding site at C-10 would be more robust compared with that at C-15, leading to the formation of the tricyclic **37**. However, the C-10 recognition site is not large enough to fully accommodate the bulkier Et group, thus mono- and bicyclic products were generated at relatively higher yields than the tricyclic product. The cyclization reactions specific to **19b**, which were not found for **20a** and **20b**, are depicted in Scheme 4D and 4E. Substrate **19b** could be converted to the tetracyclic cation **8'** through a series of ring-forming reactions: **4'** →**5'** → **6'** (like cation **6**)→**7'** (like cation **7**). The 1,2-hydride shift of H-17 to cation C-20 followed by the deprotonation of H-13 generated the tetracyclic product **31**. The tetracyclic cation **8'** underwent a normal polycyclization pathway, as shown in Scheme 1, yielding hopene homolog **33**. It is notable that the production of the tetracyclic dammarane scaffold **31** and the pentacyclic hopene skeleton **33** were not observed in the cyclization reactions of **20a** and **20b**. The placement of the bulkier Et group at C-19 of **1**, to an extent, affected the polycyclization result, leading to the generation of the tetracyclic **31** in addition to the pentacyclic product **33**, and had far-reaching effects on the reaction pathway and halted the polycyclization reactions at the premature mono- and bicyclic stages (**28**–**30** and **32**). In

 10.1002/ejoc.201800010

contrast, the substitution of the large group at C-10 and C-15 (**20a** and **20b**) completely terminated the polycyclization reaction at the abortive cyclization stages (mono- to tricycle **34**–**42**). The Et-substituent at C-19 perturbed the normal polycyclization reactions and had a strong effect on the early reaction stage(s) of the polycyclization reaction. The Me-29 binding site is not large enough to fully accommodate the Et-substituent, possibly leading to the slightly inclined geometry of **19b** around the D-ring formation site. This altered geometry may have caused inappropriate positioning of the substrate head of **19b** (A- and A/B-ring formation sites), but the hydrophobic interaction between the binding site and the Me-29 group are substantially robust, and thus, fully cyclized product **33** could be generated to some extent. Production of **29** was also limited to the reaction of **19b**. As depicted in Scheme 4E, **19b** was folded in the boat structure **43** to fulfil the product stereochemistry of the A-ring. Proton attack on the terminal double bond could yield the monocyclic cation **44** (path *k*). A 1,2-hydride shift from H-5α to C-6 cation and Me-23 shift from C-4 to C-5 followed by the deprotonation of H-3 (path *l*) gave rise to the (1,5*R*,6*R*)-trimethylcyclohexene ring **29** (IUPAC numbering). Given that **19b** is folded in a chair conformation (path *m*), the cyclization product would lead to the (1,5*R*,6*S*)-trimethylcyclohexene ring (IUPAC numbering), the stereochemistry of which is different from product **29**. The improper arrangement of the substrate head of **19b** in the reaction cavity would have a far-reaching effect on the cyclization pathway and would lead to an abnormal boat conformation during A-ring formation.

<Scheme 4>

Steric bulks of the Me-binding sites at C-6, C-10, C-15 and C-19

The yields obtained from the cyclization of **1**, **19a**, **19b**, **20a** and **20b** by the native *Aa*SHC were 100%, 0%, 18%, 12% and 9.0% (Fig. 3A), respectively, and the yield ratios were as follows: monocycle 7.2%, bicycle 7.5%, tetracycle 1.2% and pentacycle 2.4% (total yield 18%) for **19b**; monocycle 4.8%, bicycle 4.8%, tricycle 2.6% (total yield 12%) for **20a**; and monocycle 7.1% and bicycle 1.9% (total yield 9.0%) for **20b**; the absence of a reaction for **19a** indicated that the large Et-substituent at C-6 completely prevented the cyclization reaction. In contrast, a less-bulky hydrogen atom at C-6 (**11a**) afforded the fully cyclized hopene skeleton **12** without any abortive cyclization product (Scheme 2).[14] Thus, the recognition site at C-6 is noted to be compact (Fig. 3B). Substitution of the Et-group at C-19 (**19b**) afforded a broad spectrum of cyclization products, including the final product **33**, indicating that bulkiness at C-19 can be tolerated to a certain extent. However, the tolerance is low, as

evidenced by the markedly lower conversion (18%) than that of the genuine substrate **1** and by the high production yields of mono- and bicycles (see Fig. 3A). In contrast, 29-norsqualene **11b** underwent normal cyclization to yield hopane scaffolds **13** and **14** (Scheme 2A). [14] Therefore, the recognition site of position C-19 is somewhat loosely packed but is not large enough to fully accommodate the Et group. The recognition sites at positions C-10 and C-15 are more tightly packed than that at C-19, as no tetra- and/or pentacyclic product was found, and instead, only mono-, bi- and tricyclic products were found, which were produced at the early cyclization stage. Furthermore, the individual yields of **20a** and **20b** were somewhat lower than the yield of **19b**. However, the replacement of the less bulky hydrogen atom at C-15 (**12b**) afforded the normal cyclization hopene homolog **17** and the unusual folding to the boat conformation afforded **18** (opposite stereochemistry at position C-21, see Scheme 2). Thus, the Me-binding site at C-15 is not large enough to fully accommodate the Et group. The Me-recognition site at C-10 is also tightly packed but is somewhat less compact than that at C-15, as tricyclic **37** was generated as one of the enzymatic products. As described in the Introduction section, 27-norsqualene **12a** underwent abnormal conformational folding, as shown in Scheme 2A, yielding unprecedented cyclization products **15** and **16**.[15] Norsqualene **12a** lacks the Me-27 group at C-10. The recognition site at C-10 strongly captures Me-27 in **1**, but the Me-27 group is missing in norsqualene **12a**. The Me-27 recognition site incorrectly accommodates the Me-28 instead of the hydrogen atom at C-10, leading to an abnormal folding conformation, as shown in Scheme 2A. This strong hydrophobic interaction between the Me-27 binding site and the Me-27 of **1** led to the normal all-chair folding conformation shown in Scheme 1. However, the cleft for the Me-27 binding (recognition) site is not large enough to completely accommodate the Et group, and thus, the polycyclization of **20a** is halted at the tricyclic **37** stage. The cleft size of the binding site of Me-28 at C-15 is not larger than that of Me-27 of **1,** and the hydrophobic interaction is not robust; thus, the Et-substituent of **20b** could not be accommodated, and no tricyclic product was generated.

<Fig. 3>

Comparison of cyclization pathways between *Aa*SHC, and lanosterol and β-amyrin synthases

We have reported the pathways of cyclization of Et-substituted substrate analogs by hog-liver lanosterol synthase (*Hl*LAS) [20-23] and *Euphorbia tirucalli* β-amyrin synthase (*Et*AS).[16, 17, 24, 25] Fig. S14 summarizes the cyclization reactions of ethyl-substituted oxidosqualenes (OXSQs) by *Hl*LAS. The compound 10-EthylOXSQ was converted to

 10.1002/ejoc.201800010

four different products: methylenecyclohexane skeleton **43** (monocycle produced from a chair structure), cyclohexanone scaffold **44** (monocycle from a boat form), a 6,6,5-fused tricycle **45** and a 6,6,6,5-fused lanosterol homolog **46** in a ratio of ca. 1:1:1:1. In the case of the15-ethylOXSQ substrate, a 6,6,5-fused tricycle **47** and a lanosterol homolog **48** were produced in a ratio of 28:72. The compound 19-EthylOXSQ was transformed only to lanosterol homolog **49**. Notably, monocyclic products were produced only from 10-ethylOXSQ and not from 15-ethylOXSQ. In contrast, monocyclic products were generated in high yields from both 10-ethylsqualene (**20a**) and 15-ethylsqualene (**20b**) by reaction with *Aa*SHC, indicating that the Me-recognition sites at C-10 and C-15 of *Aa*SHC are more compact than those of *Hl*LAS. Furthermore, in the case of *Aa*SHC-mediated reactions, tricycle **37** was produced from 10-ethylsqualene **20a**, but little or no tricycle was obtained from the reaction of 15-ethylsqualene **20b**, indicating that the Me-binding site at C-10 of *Aa*SHC is somewhat less compact than that at C-15, unlike that of *Hl*LAS. This product pattern obtained using *Aa*SHC is clearly distinct from that obtained using *Hl*LAS. Comparison of Fig. 3B with Fig. S14.2 provides information about the difference in cleft sizes between *Aa*SHC and *Hl*LAS.

*Et*AS afforded no product from the substrates 6-, 10-, and 15-ethylOXSQs (Fig. S15.1).[17] This finding strongly suggests that the cleft sizes at the methyl-binding sites of these substrates are not large enough to accommodate the ethyl group, and thus, the ethyl-substituted OXSQs cannot be folded into a chair/chair/chair conformation in the reaction cavity. This finding indicates that the methyl-recognition sites at C-6, C-10 and C-15 of *Et*AS are tightly packed, that is, the reaction cavity size responsible for the early reaction stages (A/B/C-ring formation) is severely compact, and this tailored reaction cavity firmly folds the substrate OXSQ into a chair/chair/chair conformation.[17] In contrast, 19-ethylOXSQ afforded the β-amyrin homolog **50** (pentacycle) and tetracycle **49** in a ratio of 1:3 (Fig. S15.1) with a conversion yield of ca. 70% relative to that of the original substrate OXSQ,[24] demonstrating that the reaction cavity (the methyl-binding site at C-19) is somewhat loosely packed (Fig. S15.2), and thus, the larger ethyl group can be partially accommodated to generate the tetra- and the fully cyclized (β-amyrin) scaffolds in a ratio of 3:1.

Conclusions

In the present study conducted with *Aa*SHC, we revealed that the reaction cavities around positions C-6, C-10 and C-15 are relatively smaller than the reaction cavity at C-19, and thus, the polycyclization reactions were terminated mainly at the mono- and bicyclic stages. By comparing the polycyclization outcomes of ethyl-substituted

10.1002/ejoc.201800010

substrates using *Aa*SHC, *Hl*LAS or *Et*AS, we concluded that the promiscuity for the steric bulk at C-6, C-10 and C-15 of squalene or OXSQ increases in the following order: *Et*AS<*Aa*SHC<*Hl*LAS. Notably, the reaction cavities responsible for the early reaction stages (A/B/C-ring formation) are significantly compact, irrespective of the type of triterpene cyclase. This tightly packed reaction domain could impel each of acyclic substrates into a required folding conformation: a chair/chair/chair conformation by *Et*AS and *Aa*SHC; and a chair/boat/chair conformation by *Hl*LAS. Once these required conformations are organized in the reaction cavities around the A/B/C-ring formation site, the polyolefin cyclization occurs quickly and efficiently without the accumulation of mono-, bi- and/or tricyclic ring systems. In contrast, the reaction cavities for the later reaction stages (D/E-ring formation) are more loosely packed for all three cyclases. It is possible that this large cavity for the later reaction stages allows movement of the substrates within the reaction cavity, and thus, the large size of the cavity is likely to be one of the factors responsible for the structural diversity of the triterpene scaffolds generated; however, electronic properties, steric effects and/or hydrophobicity of the active site residues of triterpene cyclases are also crucial factors for determining the polycyclization outcomes.[2] An exceptionally large reaction domain can lead to largely unrestricted motion of OXSQ and to perturbation of either the chair or boat structure. Tetracyclic triterpene scaffolds such as lanosterol, cycloartenol and dammarane would be generated by the exceptionally loosely packing around the E-ring formation site. On the other hand, pentacyclic skeletons (hopene and plant triterpenes) would be produced by the reaction cavity at the E-ring formation site, which is relatively narrow compared to the cavity of the tetracycle-producing triterpene synthases, thus allowing folding into either the chair or boat conformation around the E-ring formation sites. We have reported the promiscuity of *Et*AS for steric sizes at C-23 of OXSQ. *Et*AS can tolerate the large ethyl group at both 23*E*-and 23*Z*-positions to afford high yields (ca 40─50% of the original substrate) of only the final products (β-amyrin skeletons), indicating that the catalytic domain around the E-ring formation site is somewhat loosely packed. [16, 25] However, *Aa*SHC-mediated reactions of squalene analogs bearing an ethyl group at C-2 (*E*- or *Z*-configured Et at initiation site) or C-23 (*E*- or *Z*-configured Et at terminal site) remain unpublished. These results will be reported in due course to provide deeper insight into the effect of the bulkiness of the squalene substrate on polycyclization by *Aa*SHC. The X-ray crystallographic structures of the two triterpene cyclases *Aa*SHC[2a, 3] and human lanosterol synthase[26] have been elucidated, and the cyclization mechanisms have been well-studied by mutation of the active site residues; [2b, 2f, 2j] however, further substrate analog experiments are required

 10.1002/ejoc.201800010

for better understanding the molecular recognition between the triterpene synthases and the substrates.

Experimental Section

Analytical Methods

NMR spectra of the enzymatic products were recorded in $CDCl_3$ or C_6D_6 on a Bruker DMX 600 and DPX 400 spectrometers, the chemical shifts being given in ppm relative to the $CDCl_3$ solvent peak δ_H =7.26 and δ_C=77.0 ppm or to the C_6D_6 solvent peak δ_H =7.28 and δ_C=128.0 ppm as the internal reference for ^{1}H- and ^{13}C NMR spectra. The coupling constants *J* are given in Hz. GC analyses were done on a Shimadzu GC2014 chromatograph equipped with a flame ionization detector (a DB-1 capillary column, 30m x 0.32 mm x 0.25 μm; J&W Scientific. Inc.). GC-MS spectra were on a JEOL SX 100 or a JEOL JMS-Q1000 GC K9 instrument equipped with a ZB-5ms capillary column (30m x 0.25 mm x 0.25 μm; Zebron) by using the EI mode operated at 70 eV. High resolution-mass spectrometry (HR-MS) was performed on a JMS-T100GCV using electron ionization (EI) mode. HPLC was carried out with Hitachi L-1700 (pump) and L-7405 (UV detector), the HPLC peaks having been monitored at 210 or 214 nm. Determination of the specific rotation values was conducted at 25°C with a Horiba SEPA-300 polarimeter (cuvette length: 10 cm) and the weights were measured with a microbalance, and the samples were dissolved in 1.0 mL of $CHCl_3$.

Syntheses of substrates 19 and 20

The experiments for the preparation of **19** are reported in the preceding paper.[17] The detailed experiments for the chemical synthesis of **20** are described in the Supporting Information, see: Fig. S1.

Incubation experiments with purified *Aa*SHC and the cell-free extract

The *Escherichia coli* (DE3) encoding pET26b-SHC and pET3a-SHC was cultured under the conditions reported in the previous papers.[18, 19] The detailed incubation conditions are described in the Text (see Fig. 1).

Spectroscopic data of products 28─41.

EIMS and the ^{1}H- and ^{13}C-NMR spectra including 2D NMR and the assignments of ^{1}H- and ^{13}C-signals are described in the Supporting information Figs. S2─S13. Products **28**, **29**, **30**, **31**, **32**, **33**, **37** and **41** were isolated in a pure state, thus the complete assignments of ^{1}H- and ^{13}C-NMR signals of these products were determined

 10.1002/ejoc.201800010

as follows.

Product 28. $[\alpha]_D^{25}$ = +18 (*c* = 0.27, $CHCl_3$). ^{1}H NMR (400 MHz, $CDCl_3$): δ=0.833 (s, 3H, 24-Me), 0.911 (s, 3H, 23-Me), 0.959 (t, 3H, *J*=7.6 Hz, Me-31), 1.22 (m, 1H, 3-H), 1.43 (m, 1H, 7-H), 1.47 (m, 1H, 3-H), 1.53 (m, 2H, 2-H), 1.56 (m, 1H, 7-H), 1.591 (s, 3H, 26-Me), 1.606 (s, 6H, 27-Me and 30-Me), 1.683 (s, 3H, 29-Me), 1.69 (m, 1H, 5-H), 1.75 (m, 1H, 8-H), 1.94 (m, 1H, 8-H), 1.98 (m, 2H, 19-H), 2.00 (m, 3H, 1-H and 15-H), 2.02 (m, 6H, 11-H, 16-H and 20-H), 2.03 (m, 2H, 28-H), 2.05 (m, 1H, 1-H), 2.08 (m, 2H, 12-H), 4.54 (s, 1H, 25-H), 4.75 (s, 1H, 25-H), 5.15-5.06 (m, 4H, 10-H, 13-H, 17-H, 21-H) ppm. The assignment of 11-H and 12-H may be exchangeable. ^{13}C NMR (100 MHz, $CDCl_3$): δ=13.2 (q, C-31), 16.1 (q, 2C, C-26 and C-27), 17.7 (q, C-30), 23.2 (t, C-28), 23.8 (t, C-2), 24.8 (t, C-7), 25.7 (q, C-29), 26.2 (q, C-24), 26.4 (t, C-12), 27.0 (t, C-11), 28.3 (t, 2C, C-16 and C-20), 28.4 (q, C-23), 32.5 (t, C-1), 34.9 (s, C-4), 36.4 (t, C-3), 36.5 (t, C-19), 38.2 (t, C-8), 40.1 (t, C-15), 53.6 (d, C-5), 108.8 (t, C-25), 123.8 (d, C-17), 124.0 (d, C-10), 124.3 (d, C-21), 124.5 (d, C-13), 131.2 (s, C-22), 135.1 (s, C-14), 135.7 (s, C-9), 140.8 (s, C-18), 149.4 (s, C-6) ppm. The assignment of C-11 and C-12 and that of C-13 and C-17 may be exchangeable. MS (EI) see Fig. S2.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4073.

Product 29. $[\alpha]_D^{25}$ = - 6 (*c* = 0.09, $CHCl_3$). ^{1}H NMR (600 MHz, C_6D_6): δ=0.979 (d, 3H, *J*=6.8 Hz, Me-25), 0.979 (t, 3H, *J*=7.5 Hz), 1.01 (s, 3H, 24-Me), 1.53 (m, 2H, 1-H) 1.698 (s, 3H, 30-Me), 1.70 (m, 2H, 7-H), 1.747 (s, 3H, 27-Me), 1.766 (s, 3H, 26-Me), 1.791(s, 3H, 23-Me), 1.810 (s, 3H, 29-Me), 1.87 (m, 1H, 6-H), 1.93 (m, 8-H), 2.03 (m, 1H, 2-H), 2.10 (m, 1H, 2-H), 2.16 (mm 8-H), 2.20 (m, 2H, 28-H), 2.23 (m, 2H, 15-H), 2.25 (m, 2H, 19-H), 2.28 (m, 4H, 16-H and 20-H), 2.32 (m, 2H, 11-H), 2.35 (m, 2H, 12-H), 5.38 (t, 2H, *J*=6.7 Hz, 17-H and 21-H), 5.47 (s, 2H, 10-H and 13-H), 5.59 (s, 1H, 3-H) ppm. The assignments of 11-H and 12-H may be exchangeable. ^{13}C NMR (150 MHz, C_6D_6): δ=13.5 (q, C-31), 16.0 (q, C-25), 16.2 (q, C-27), 16.4 (q, C-26), 17.7 (q, C-30), 19.4 (q, C-23), 21.2 (q, C-24), 23.6 (t, C-28), 25.8 (q, C-29), 25.9 (t, C-2), 26.8 (t, C-12), 27.4 (t, C-11), 27.5 (t, C-1), 28.8 (t, 2C, C-16 and C-20), 33.5 (d, C-6), 34.7 (t, C-8), 35.7 (t, C-7), 37.0 (t, C-19), 40.6 (t, C-15), 40.7 (s, C-5), 124.1 (d, C-10), 124.4 (d, C-17), 124.5 (t, C-3), 124.9 (d, C-13), 125.0 (d, C-21), 131.0 (s, C-22), 135.2 (s, C-14), 136.2 (s, C-9), 139.7 (s, C-4), 140.9 (s, C-18) ppm. The assignments of C-1, C-11 and C-12 may be exchangeable. MS (EI): see Fig. S3.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4054.

Product 30. $[\alpha]_D^{25}$ = +25 (*c* = 0.186, $CHCl_3$). ^{1}H NMR (400 MHz, $CDCl_3$): δ=0.666 (s, 3H, 25-Me), 0.800 (s, 3H, 24-Me), 0.868 (s, 3H, 23-Me), 0.963 (t, 3H, *J*=7.6 Hz, 31-Me), 1.00 (m, 1-H), 1.08 (dd, 1H, *J*=12.8, 2.8Hz, 5-H), 1.18 (ddd, 1H, *J*=13.0, 13.0, 4.1

 10.1002/ejoc.201800010

Hz, H-3), 1.32 (m, 1H, 6-H), 1.37 (m, 2H, 2-H and11-H), 1.38 (m, 1H, 3-H), 1.48 (m, 1H, 11-H), 1.567 (s, 3H, 27-Me), 1.57 (m, 2-H), 1.604 (s, 3H, 30-Me), 1.61 (m, 1H, 9-H), 1.64 (m, 1-H), 1.683 (s, 3H, 29-Me), 1.71 (m, 1H, 6-H), 1.82 (m, 1H, 12-H), 1.97 (m, 1H, 7-H), 1.99 (m, 4H, 15-H and 19-H), 2.03 (m, 2H, 28-H), 2.04 (m, 4H, 16-H and 20-H), 2.08 (m, 1H, 12-H), 2.38 (m, 1H, 7-H), 4.54 (s, 1H, 26-H), 4.82 (s, 1H, 26-H), 5.08 (m, 1H, 17-H), 5.12 (m, 1H, 21-H), 5.13 (m, 1H, 13-H) ppm. ^{13}C NMR (100 MHz, $CDCl_3$): δ=13.3 (q, C-31), 14.5 (q, C-25), 16.0 (q, C-27), 17.7 (q, C-30), 19.4 (t, C-2), 21.7 (q, C-24), 23.2 (t, C-28), 23.8 (t, C-11), 24.5 (t, C-6), 25.7 (q, C-29), 26.4 (t, C-20) 26.9 (t, C-16), 27.0 (t, C-12), 33.6 (q, 2C, C-4 and C-23), 36.6 (t, C-19), 38.4 (t, C-7), 39.1 (t, C-1), 39.6 (s, C-10), 40.1 (t, C-15), 42.2 (t, C-3), 55.6 (d, C-5), 56.2 (d, C-9), 106.1 (t, C-26), 123.9 (d, C-17), 124.5 (d, C-21), 125.1 (d, C-13), 131.2 (s, C-22), 134.9 (s, C-14), 140.8 (s, C-18), 148.8 (s, C-8) ppm. The assignments of C-12, C-16 and C-200 may be exchangeable. MS (EI): see Fig. S4.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4060.

Product **31**. $[\alpha]_D^{25}$ = - 41 (*c* = 0.11, $CHCl_3$). ^{1}H NMR (600 MHz, C_6D_6): δ=0.91 (m, 1H, 1-H), 0.940 (m, 1H, 5-H), 0.959 (t, 3H, *J*=7.1 Hz, Me-31), 0.972 (s, 3H, 29-Me), 0.997 (s, 3H, 19-Me), 1.038 (s, 3H, 28-Me), 1.122 (s, 3H, 18-Me), 1.29 (m, 1H, 3-H), 1.312 (s, 3H, 30-Me), 1.40 (m, 1H, 11-H), 1.42 (m, 1H, 6-H), 1.44 (m, 1H, 15-H), 1.46 (m, 1H, 21-H), 1.48 (m, 1H, 7-H), 1.50 (m, 1H, 3-H), 1.51 (m, 1H, 21-H), 1.52 (m, 1H, 6-H), 1.58 (m, 2H, 22-H), 1.60 (m, 1H, 9-H), 1.66 (m, 1H, 2-H), 1.67 (m, 1H, 7-H), 1.68 (m, 1H, 11-H), 1.73 (m, 1H, 2-H), 1.741 (s, 3H, 27-Me), 1.78 (m, 1H, 1-H), 1.839 (s, 3H, 26-Me), 2.03 (m, 1H, 12-H), 2.08 (m, 1H, 23-H), 2.11 (m, 1H, 15-H), 2.27 (m, 1H, 16-H), 2.30 (m, 1H, 23-H), 2.35 (m, 1H, 16-H), 2.42 (m, 1H, 20-H), 2.59 (dd, 1H, *J*=14.0, 3.4 Hz, 12-H), 5.44 (bs, 1H, 24-H) ppm. ^{13}C NMR (150 MHz, C_6D_6): δ=12.6 (q, C-31),

16.6 (q, C-19), 17.8 (q, 2C, C-18 and C-27), 18.9 (t, C-6), 19.1 (t, C-2), 21.9 (q, C-29), 22.0 (t, C-11), 23.2 (q, C-30), 23.6 (t, C-12), 25.9 (q, C-26), 27.1 (t, C-23), 27.7 (t, C-21), 29.1 (t, C-16), 30.9 (t, C-15), 33.5 (s, C-4), 34.5 (t, C-22), 35.8 (t, C-7), 38.0 (s, C-10), 38.7 (q, C-28), 39.8 (d, C-20), 40.9 (t, C-1), 41.3 (s, C-8), 42.3 (t, C-3), 52.1 (d, C-9), 57.1 (s, C-14), 57.3 (d, C-5), 125.6 (d, C-24), 130.8 (s, C-25), 132.7 (s, C-17), 142.0 (s, C-13) ppm. The assignments of C-2 and C-6 and those of C-18 and C-27 may be exchangeable. MS (EI): see Fig. S5.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4083.

Product 32. $[\alpha]_D^{25}$ = +12 (*c* = 0.11, $CHCl_3$). ^{1}H NMR (400 MHz, $CDCl_3$): δ=0.741 (s, 3H, 25-Me), 0.852 (s, 3H, 23-Me), 0.874 (s, 3H, 24-Me), 0.96 (m, 1H, 1-H), 0.980 (t, *J*=6.8 Hz, 3H, 31-Me), 1.16 (m, 1H, 3-H), 1.17 (dd, *J*=12.8, 4.8Hz, 1H, 5-H), 1.21 (m,

1H, 11-H), 1.29 (m, 1H, 3-H), 1.45 (m, 2H, 2-H and 11-H), 1.51 (m, 1H, 2-H), 1.604 (s, 3H, 30-Me), 1.614 (s, 3H, 27-Me), 1.63 (m, 1H, 9-H), 1.683 (s, 3H, 29-Me), 1.710 (bs, 3H, 26-Me), 1.84 (m, 1H, 1-H), 1.85 (m, 1H, 6-H), 1.94 (m, 1H, 6-H), 1.97 (m, 1H, 12-H), 2.00 (m, 4H, 15-H and 19-H), 2.03 (m, 4H, 16-H and 28-H), 2.08 (m, 2H, 20-H) 2.17 (m, 1H, 12-H), 5.08 (m, 1H, 17-H), 5.13 (m, 2H, 13-H and 21-H), 5.38 (bs, 1H, 7-H) ppm. The assignments of 16-H and 20-H may be exchangeable. ^{13}C NMR (100 MHz, $CDCl_3$): δ=13.3 (q, C-31), 13.6 (q, C-25), 16.2 (q, C-27), 17.7 (q, C-30), 18.8 (t, C-2), 21.9 (q, C-24), 22.2 (q, C-26), 23.2 (t, C-28), 23.8 (t, C-6), 25.7 (q, C-29), 26.4 (t, C-20), 27.3 (t, 2C, C-11 and C-16), 30.2 (t, C-12), 33.0 (s, C-4), 33.2 (q, C-23), 36.6 (t, C-19), 36.7 (s, C-10), 39.2 (t, C-1), 40.1 (t, C-15), 42.4 (t, C-3), 50.2 (d, C-5), 54.2 (d, C-9), 122.0 (d, C-7), 123.8 (d, C-17), 124.3 (d, C-13), 124.5 (d, C-21), 131.2 (s, C-22), 135.0 (s, C-14), 135.6 (s, C-8), 140.9 (s, C-18) ppm. The assignments of C-16 and C-20 may be exchangeable. MS (EI): see Fig. S6.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4039.

Product 33. $[\alpha]_D^{25}$ = + 16 (*c* = 0.121, $CHCl_3$). ^{1}H NMR (600 MHz, C_6D_6): δ=0.87 (bd, *J*=11.4 Hz, 1H, 5-H), 0.89 (m, 1H, 1-H), 0.93 (m, 1H, 19-H), 0.977 (s, 3H, 25-Me) 0.997 (s, 3H, 24-Me), 1.021 (t, *J*=7.4 Hz, 3H, 31-Me), 1.046 (s, 3H, 23-Me), 1.101 (s, 3H, 26-Me), 1.187 (s, 3H, 27-Me), 1.29 (m, 1H, 3-H), 1.32 (m, 1H, 7-H), 1.35 (m, 2H, 11-H and 15-H), 1.38 (m, 1H, 9-H), 1.45 (m, 1H, 17-H), 1.46 (m, 2H, 6-H and 15-H), 1.50 (m, 2H, 3-H and 13-H), 1.52 (m, 1H, 6-H), 1.57 (m, 1H, 28-H), 1.62 (m, 2H, 7-H and 11-H), 1.64 (m, 2H, 2-H and 16-H), 1.71 (m, 2H, 12-H), 1.73 (m, 1H, 2-H), 1.76 (m, 1H, 16-H), 1.77 (m, 2H, 1-H and 19-H), 1.80 (m, 1H, 28-H), 1.87 (m, 1H, 20-H) 1.912 (s, 3H, 30-Me), 1.98 (m, 1H, 20-H), 2.72 (m, 1H, 21-H), 5.07 (s, 1H, 29-H), 5.11 (s, 1H, 29-H) ppm. ^{13}C NMR (150 MHz, C_6D_6): δ=10.6 (q, C-31), 16.1 (q, C-25) 17.3 (q, C-26), 17.6 (q, C-27), 19.1 (t, 2C, C-2 and C-6), 20.0 (t, C-28), 21.5 (t, C-16), 21.8 (q, C-24 and t, C-11), 25.7 (q, C-30), 26.5 (t, C-12), 28.1 (t, C-20), 33.4 (s, C-4), 33.6 (t, C-7 and q, C-23), 34.3 (t, C-15), 36.9 (t, C-19), 37.6 (s, C-10), 40.6 (t, C-1), 42.4 (t, C-3), 42.5 (s, 2C, C-8 and C-14), 46.3 (d, C-21), 47.4 (s, C-18), 50.9 (d, C-9), 51.3 (d, C-13), 56.3 (d, C-17), 56.5 (d, C-5), 109.9 (t, C-29), 148.9 (s, C-22) ppm. MS (EI): see Fig. S7.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4061.

Product 37. $[\alpha]_D^{25}$ = + 49 (*c* = 0.06, $CHCl_3$). ^{1}H NMR (600 MHz, $CDCl_3$): δ=0.734 (s, 3H, 25-Me), 0.772 (t, *J*=7.4Hz, 31-Me), 0.823 (d, *J*=6.7Hz, 27-Me), 0.869 (s, 3H, 23-Me), 0.905 (s, 3H, 24-Me), 0.91 (m, 1H, 15-H), 1.05 (m, 1H, 1-H), 1.18 (m, 1H, 3-H), 1.20 (m, 1H, 11-H), 1.22 (m, 1H, 5-H), 1.33 (m, 1H, 12-H), 1.35 (m, 1H, 26-H), 1.38 (m, 1H, 14-H), 1.40 (m, 1H, 2-H), 1.42 (m, 1H, 3-H), 1.48 (m, 1H, 11-H), 1.51 (m, 2H,

 10.1002/ejoc.201800010

12-H and 15-H), 1.52 (m, 1H, 26-H), 1.55 (m, 1H, 2-H), 1.599 (s, 3H, 28-Me), 1.60 (m, 1H, 1-H), 1.604 (s, 3H, 30-Me), 1.683 (s, 3H, 29-Me), 1.85 (m, 1H, 16-H), 1.88 (m, 1H, 6-H), 1.95 (m, 1H, 9-H), 1.97 (m, 2H, 19-H), 2.06 (m, 1H, 16-H), 2.07 (m, 2H, 20-H), 2.08 (m, 1H, 6-H), 5.11 (m, 2H, 17-H and 21-H), 5.20 (bs, 1H, 7-H) ppm. ^{13}C NMR (150 MHz, $CDCl_3$): δ=9.11 (q, C-31), 13.6 (q, C-25), 14.6 (q, C-27), 16.0 (q, C-28), 17.7 (q, C-30), 18.9 (t, C-2), 21.7 (q, C-24), 23.6 (t, C-6), 24.2 (t, C-11), 25.7 (q, C-29), 26.7 (t, C-16), 26.8 (t, C-20), 30.4 (t, C-26), 31.9 (t, C-12), 31.9 (t, C-15), 32.9 (s, C-4), 33.4 (q, C-23), 34.9 (s, C-10), 39.8 (t, C-19), 39.9 (d, C-14), 40.4 (t, C-1), 42.8 (t, C-3), 50.7 (d, C-5), 50.7 (s, C-13), 58.8 (d, C-9), 115.7 (d, C-7), 124.5 (d, C-21), 125.1 (d, C-17), 131.2 (s, C-22), 134.6 (s, C-18), 147.1 (s, C-8) ppm. The assignments of C-12 and C-15 and those of C-16 and C-20 may be exchangeable. MS (EI): see Fig. S10.1 in the Supporting Information. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4059.

Product 41. $[\alpha]_D^{25}$ = + 46 (*c* = 0.02, $CHCl_3$). ^{1}H NMR (600 MHz, C_6D_6): δ=0.872 (s, 3H, 24-Me), 1.056 (s, 3H, 23-Me), 1.122 (t, *J*=7.4 Hz, 3H, 31-Me), 1.202 (s, 3H, 25-Me), 1.21 (m, 1H, 3-H), 1.22 (m, 1H, 5-H), 1.30 (ddd, *J*=12.2,12.2, 3.7 Hz, 1H, 1-H), 1.36 (m, 1H, 3-H), 1.42 (m, 1H, 2-H), 1.52 (m, 1H, 2-H), 1.56 (m, 1H, 7-H), 1.694 (s, 3H, 30-Me), 1.72 (m, 1H, 1-H), 1.745 (s, 3H, 28-Me), 1.808 (s, 3H, 29-Me), 1.848 (s, 3H, 26-Me), 1.85 (m, 1H, 7-H), 2.19 (q, *J*=7.4Hz, 2H, 27-H), 2.23 (m, 2H, 19-H), 2.26 (m, 2H, 15-H), 2.30 (m, 4H, 11-H and 12-H), 2.30 (m, 4H, 16-H and 20-H), 2.33 (m, 1H, 8-H), 2.46 (m, 1H, 8-H), 5.38 (t, *J*=6.9 Hz, 1H, 21-H), 5.43 (m, 1H, 17-H), 5.45 (m, 1H, 13-H), 5.54 (bs, 10-H) ppm. The assignments of 13-H and 17-H may be exchangeable. ^{13}C NMR (150 MHz, C_6D_6): δ=13.4 (q, C-31), 16.1 (q, C-28), 16.3 (q, C-26), 17.7 (q, C-30), 20.8 (t, C-2), 21.6 (q, C-24), 23.6 (q, 2C, C-25 and C-27), 25.4 (t, C-7), 25.8 (q, C-29), 27.2 (t, C-20), 27.4 (t, C-16), 28.4 (t, C-11), 29.0 (t, C-12), 32.9 (q, C-23), 35.6 (s, C-4), 37.0 (t, C-15), 40.2 (t, C-19), 41.8 (t, C-3), 43.4 (t, C-8), 43.95 (t, C-1), 56.9 (d, C-5), 73.5 (s, C-6), 124.5 (d, C-13), 124.6 (d, C-10), 124.9 (d, 2C, C-17 and C-21), 131.0 (s, C-22), 134.9 (s, C-18), 136.5 (s, C-9), 141.1 (s, C-14) ppm. The assignments of C-11 and C-12 and those of C-16 and C-20 may be exchangeable. MS (EI): see Fig. S12.1 in the Supporting Information; the molecular ion *m*/*z* 442 (M^+) was not observed, but *m*/*z* 424 (M^+-H_2O) was clearly observed. HRMS (EI): *m*/*z*: calcd. 424.4069 for $C_{31}H_{52}$, found: 424.4065.

Other products. As described in Text, the separation of a mixture of **34** and **35**, that of **36** and **38**, and that of **39** and **40** failed. Therefore, the NMR spectra were provided as a mixture of the two products. The chemical shifts of the ^{1}H- and ^{13}C-signals, which were differentiated between **34** and **35**, between **36** and **38** and between **39** and **40**, are described in Fig. S8.9, Fig. S9.9 and Fig. S11.9, respectively.

10.1002/ejoc.201800010

Acknowledgements

This work was supported in part by Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science, Nos. 25450150 (C) and 18380001(B).

Keywords: Alkene; cyclization, enzyme catalysis; polycycle; terpenoids

10.1002/ejoc.201800010

References

[1] a) R. B. Woodward and K. Bloch, *J. Am. Chem. Soc.*, **1953**, *75*, 2023-2024. b) A. Eschenmoser, L. Ruzicka, O. Jeger and D. Arigoni, *Helv. Chim. Acta* **1955**, *38*, 1890–1904. c) J. W. Cornforth, R. H. Cornforth and C. Donniger, G. Popjak, Y. Shimizu, S. Ichii, E. Forchielli and E. Caspi, *J. Am. Chem. Soc.*, **1965**, *87*, 3224-3228.

[2] For the reviews of squalene and oxidosqualene cyclases, see: a) K. U. Wendt, G. E. Schulz, E. J. Corey and D. R. Liu, *Angew. Chem., Int. Ed.* **2000**, *39*, 2812–2833. b) T. Hoshino and T. Sato, *Chem. Commun.* **2002**, 291–301. c) R. A. Yoder and J. N. Johnston, *Chem. Rev.* **2005**, *105*, 4730–4756. d) M. J. R. Segura, B. E. Jackson and S. P. T. Matsuda, *Nat. Prod. Rep.*, **2003**, *20*, 304-317. e) I. Abe, *Nat. Prod. Rep.* **2007**, *24*, 1311–1331. f) T. K. Wu, C. H. Chang, Y. T. Liu and T. T. Wang, *Chem. Rec.* **2008**, *8*, 302–325. g) W. D. Nes, *Chem. Rev.* **2011**, *111*, 6423–6451. h) G. Siedenburg, D. Jendrossek, *Appl. Environ. Microbiol.*, **2011**, *77*, 3905-3915. i) P.-O. Syrén, S. Henche, A. Eichler, B. M. Nestl, B. Hauer, *Curr. Opin. Struct. Biol.* **2016**, *41*, 73-82. j) T. Hoshino, *Org. Biomol. Chem.*, **2017**, *15*, 2869-2891.

[3] K. U. Wendt, K. Poralla and G. E. Schulz, *Science*, **1997**, *277*, 1811-1815.

[4] (a) T. Sato and T. Hoshino, *Biosci., Biotechnol., Biochem.*, **1999**, *63*, 2189–2198; (b) T. Dang and G. D. Prestwich, *Chem. Biol.*, **2000**, *7*, 643–649.

[5] T. Hoshino and T. Sato, *Chem. Commun.*, **1999**, 2005–2006.

[6] C. Pale-Grosdemange, T. Merkofer, M. Rohmer and K. Poralla, *Tetrahedron Lett.*, **1999**, *40*, 6009–6012.

[7] T. Hoshino, M. Kouda, T. Abe and S. Ohashi, *Biosci. Biotechnol. Biochem.*, **1999**, *63*, 2038–2041.

[8] C. Pale-Grosdemange, C. Feil, M. Rohmer and K. Poralla, *Angew. Chem., Int. Ed.*, **1998**, *37*, 2237–2240.

[9] T. Sato, T. Abe and T. Hoshino, *Chem. Commun.*, **1998**, 2617–2618.

[10] T. Merkofer, C. Pale-Grosdemange, K. U. Wendt, M. Rohmer and K. Poralla, *Tetrahedron Lett.*, **1999**, *40*, 2121–2124.

[11] T. Hoshino, T. Abe and M. Kouda, *Chem. Commun.*, **2000**, 441–442.

[12] (*a*) T. Hoshino, M. Kouda, T. Abe and T. Sato, *Chem. Commun.*, **2000**, 1485–1486; (*b*) Related paper: S. Schmitz, C. Füll, T. Glaser, K. Albert and K. Poralla, *Tetrahedron Lett.*, **2001**, 42, 883–885.

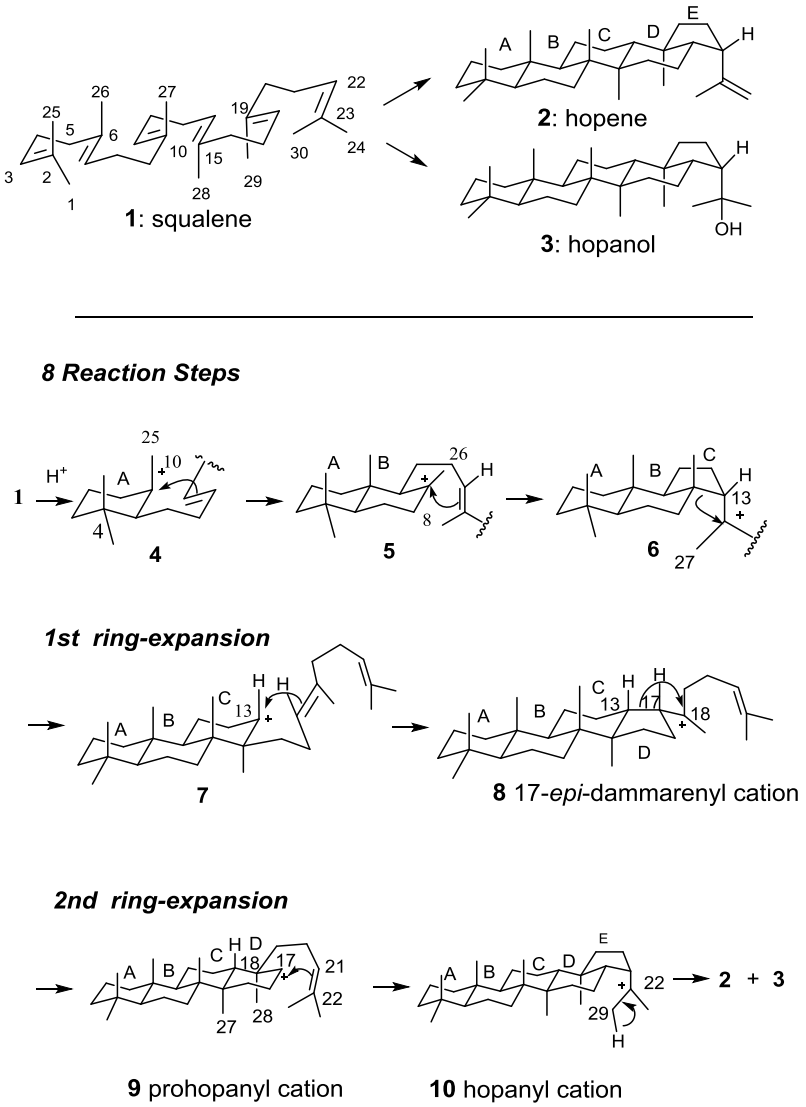
[13] T. Hoshino, S. Nakano, T. Kondo, T. Sato and A. Miyoshi, *Org. Biomol. Chem.*, **2004**, *2*, 1456-1470.

10.1002/ejoc.201800010

[14] S. Nakano, S. Ohashi and T. Hoshino, *Org. Biomol. Chem.*, **2004**, *2*, 2012-2022.
[15] T. Hoshino and S. Ohashi , *Org. Lett.*, **2002**, *4*, 2553-2556.
[16] T. Hoshino, Y. Miyahara, M. Hanaoka, K. Takahashi and I. Kaneko, *Chem. Eur. J.*, **2015**, *21*, 15769 – 15784.
[17] Y. Terasawa, Y. Sasaki, Y. Yamaguchi, K. Takahashi and T. Hoshino, *Eur. J. Org. Chem.*, **2017**, 287-295.
[18] M. Morikubo, Y. Fukuda, K. Ohtake, N. Shinya, D. Kiga, K. Sakamoto, M. Asanuma, H. Hirota, S. Yokoyama and T. Hoshino, *J. Am. Chem. Soc.*, **2006**, *128*, 13148-13194.
[19] a) T. Sato, Y. Kanai and T. Hoshino, *Biosci. Biotechnol. Biochem.*, **1998**, *62*, 407-411. b) T. Sato and T. Hoshino, *Biosci. Biotechnol. Biochem.*, **1999**, *63*, 2189-2198. c) T. Sato, S. Sasahara, T. Yamakami and T. Hoshino, *Biosci. Biotechnol. Biochem.*, **2002**, *66*, 1660-1670.
[20] T. Hoshino and Y. Sakai, *Tetrahedron Lett.*, **2001**, *42*, 7319-7323.
[21] T. Hoshino, E. Ishibashi, *J. Chem. Soc., Chem. Commun.*, **1995**, 2401-2402.
[22] T. Hoshino and Y. Sakai, *Chem. Commun.*, **1998**, 1591-1592.
[23] T. Hoshino, A. Chiba and N. Abe, *Chem. Eur. J.*, **2012**, *18*, 13108-13116.
[24] T. Hoshino, Y. Yamaguchi, K. Takahashi and R. Ito, *Org. Lett.*, **2014**, *16*, 3548-3551.
[25] I. Kaneko and T. Hoshino, *J. Org. Chem.*, **2016**, *81*, 6657-6671.
[26] R. Tohma, T. Schulz-Gasch, B. D'Archy, J. Benz, J. Arbi, H. Dehmlow, M. Henning, M. Stihle and A. Ruf, *Nature*, **2004**, *432*,118-122.

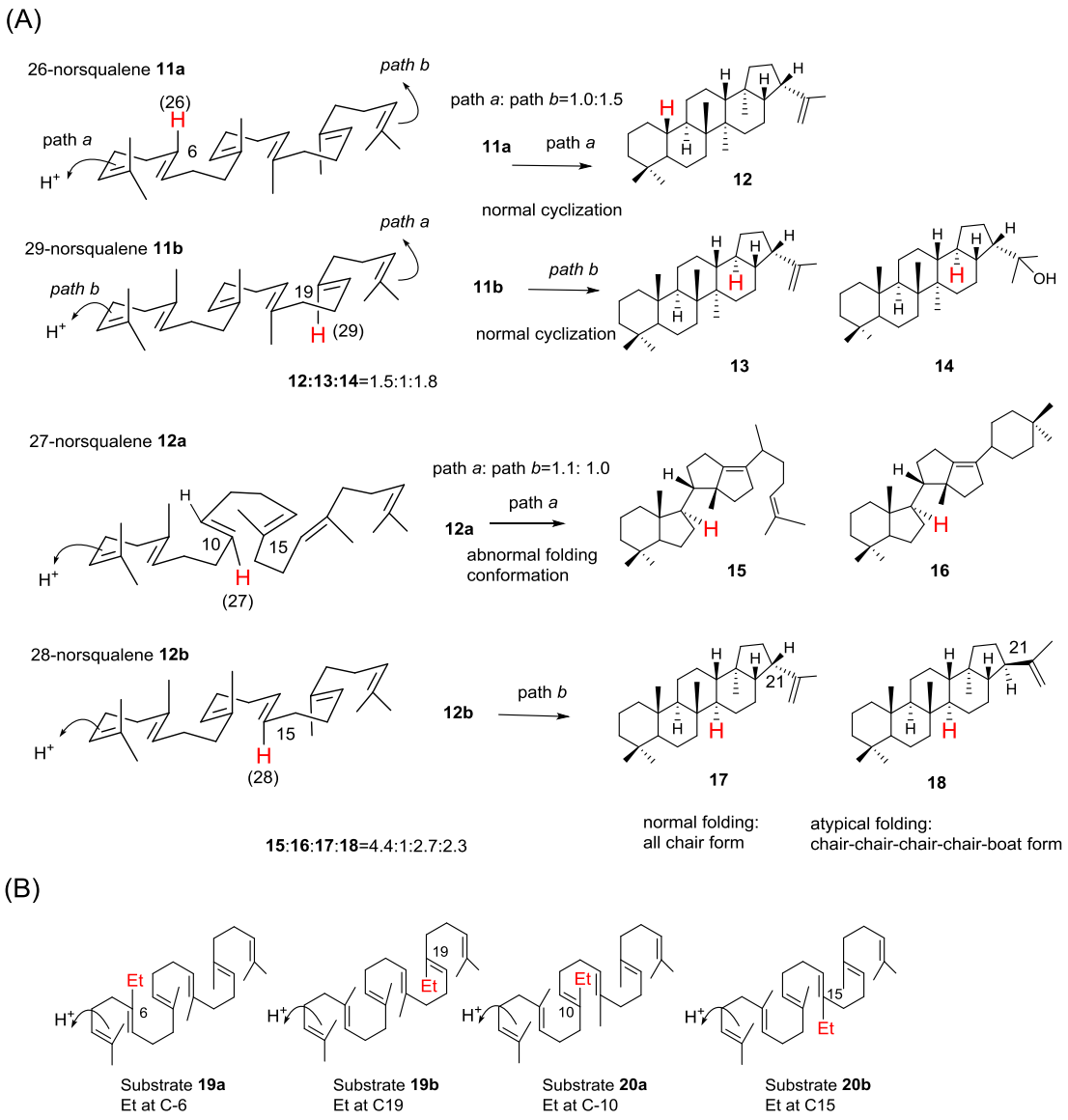
10.1002/ejoc.201800010

Figures and Schemes



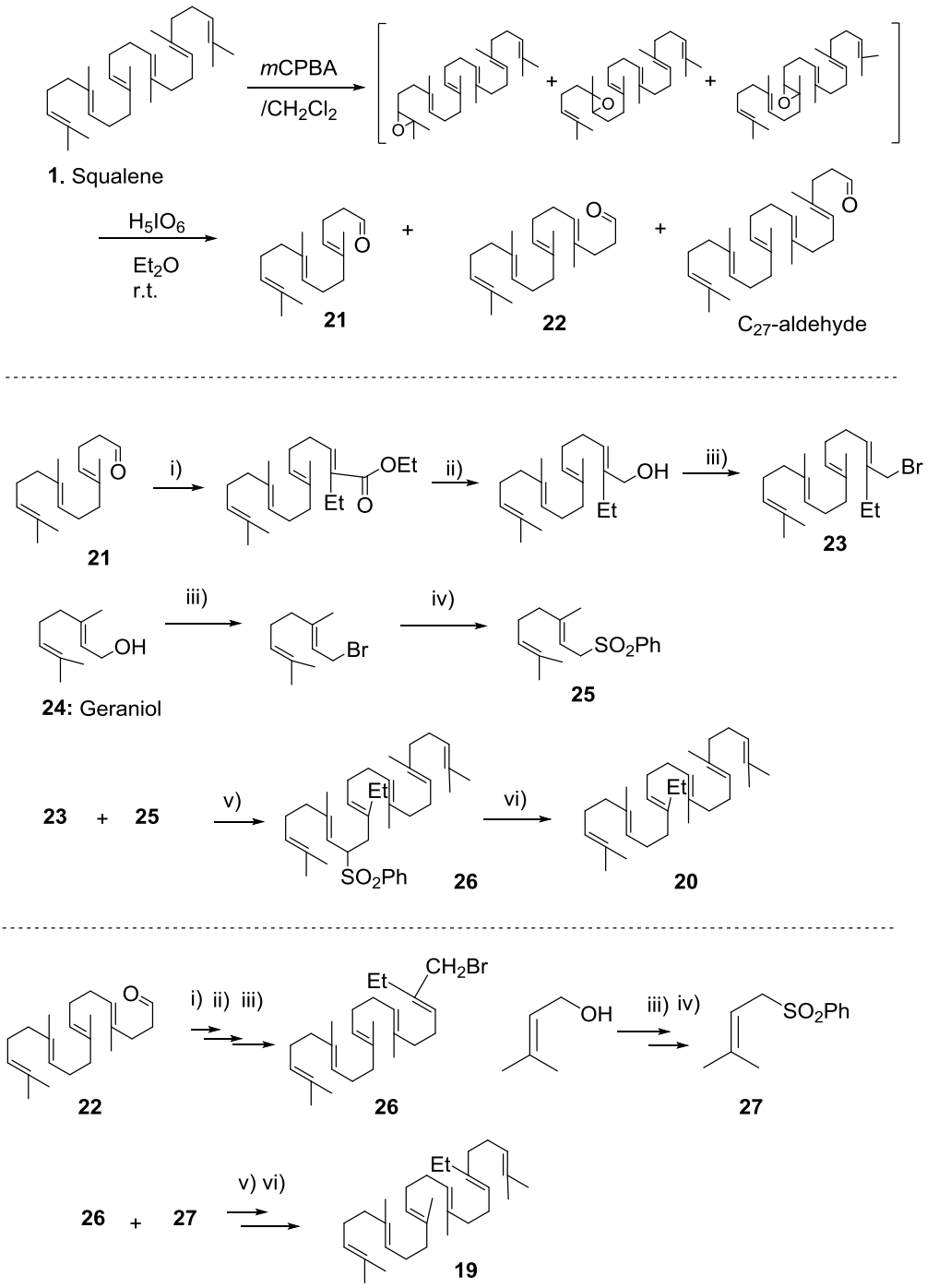
Scheme 1. Polycyclization pathway of squalene **1** to pentacyclic-hopene **2** and hopanol **3** proposed based on the abortive cyclization products afforded by the site-directed variants.

10.1002/ejoc.201800010



Scheme 2. (A) Cyclization pathway of norsqualene analogs **11** and **12**. Two pathways *a* and *b* are possible, as the isopropylidene moieties, which are necessary for initiation of the cyclization reaction,[13} are present at the both termini. (B) Structures of Et-substituted **19** and **20**. Two names are proposed for both **19** and **20** based on the two different directions for initiating the polycyclization reactions: 6-ethylsqualene **19a**, 19-ethylsqualene **19b**, 10-ethylsqualene **20a** and 15-ethylsqualene **20b**.

10.1002/ejoc.201800010



i) triethyl-2-phosphonobutyrate $(EtO)_2P(O)CH(Et)CO_2Et$/*n*-BuLi,rt; ii) $^iBu_2AlH/Et_2O$, -40°C; iii) PBr_3/ Et_2O, 0°C; iv) $PhSO_2Na$/DMF, rt; v) *n*-BuLi/THF-HMPA (9:1), -78°C; vi) $LiBEt_3H/Et_2O$, $[PdCl_3(dppp)]$.

Scheme 3. Schemes for the syntheses of ethylated analogs **19** and **20**.

10.1002/ejoc.201800010

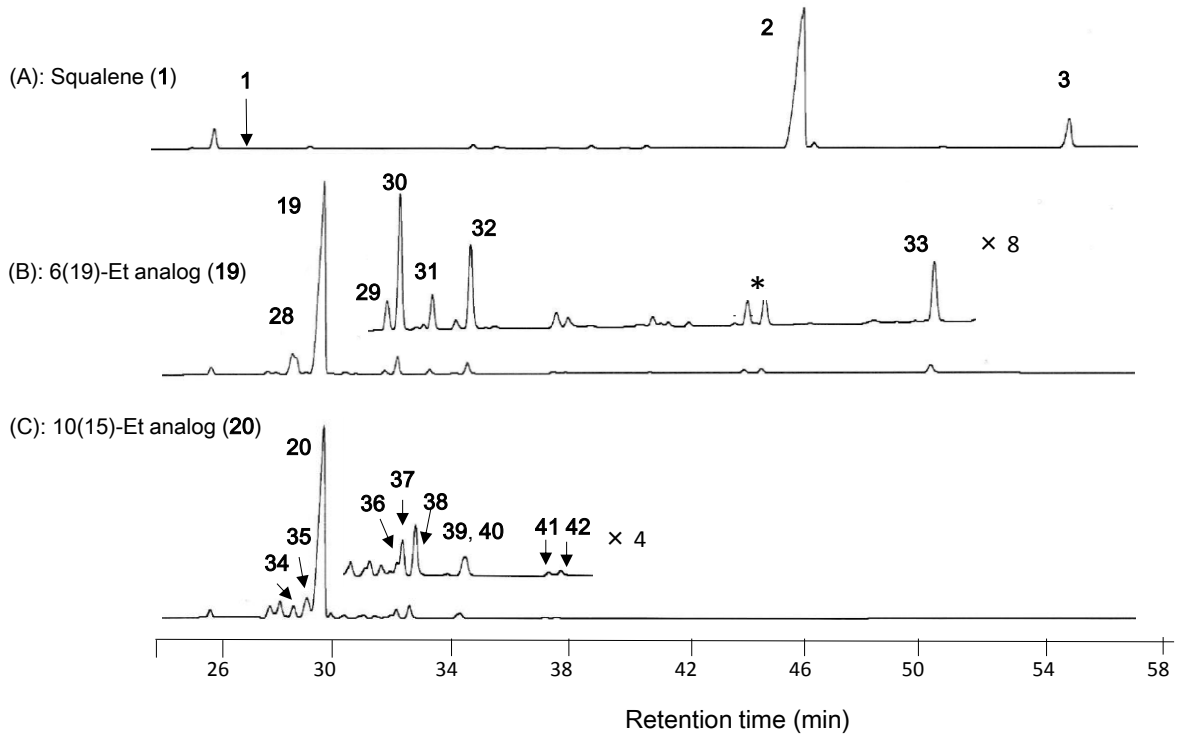
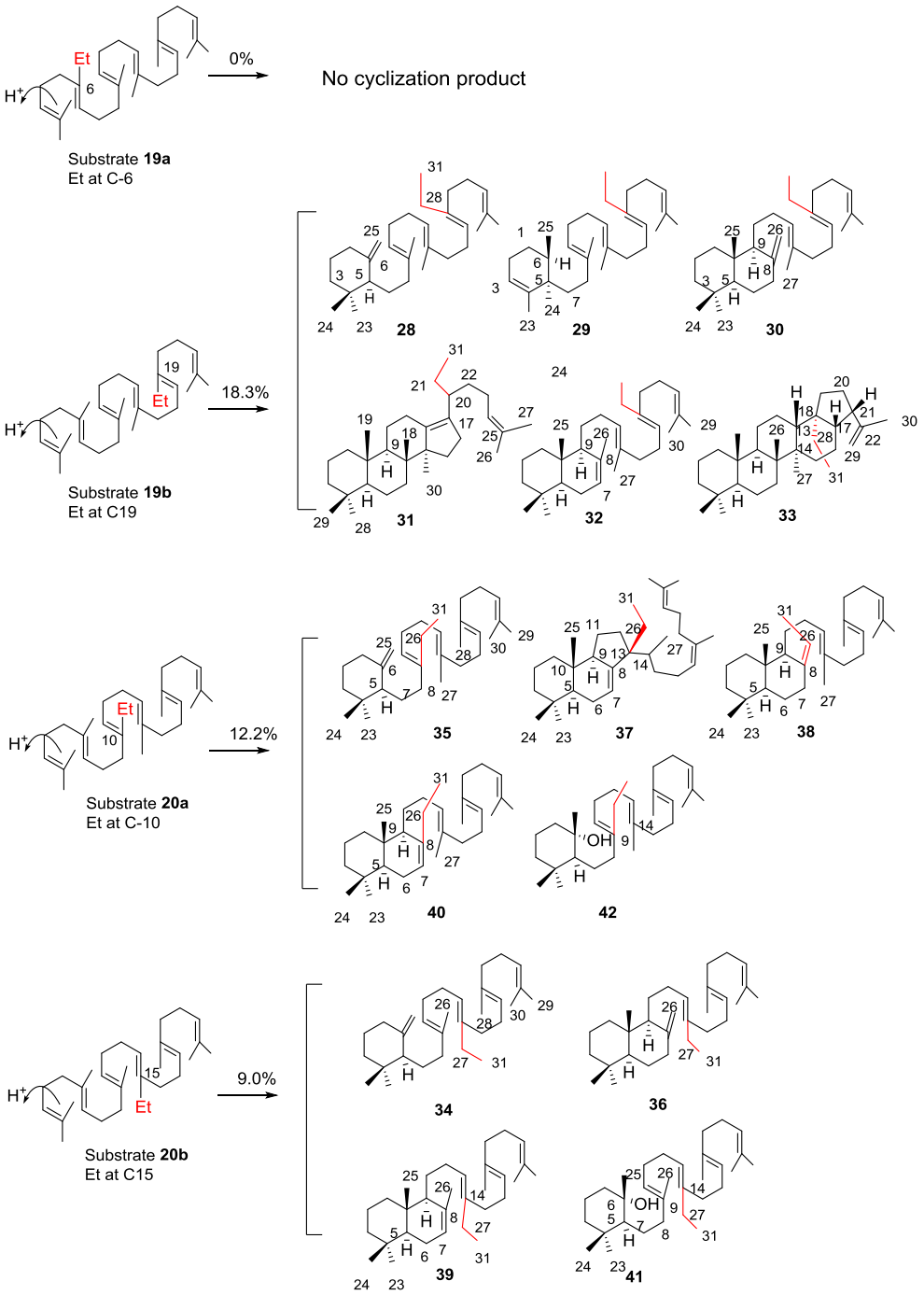


Fig. 1. Gas chromatograms of the reaction mixture obtained by separately incubating **1** (A), **19** (B) and **20** (C) with wild-type *Aa*SHC. Triton X-100 was removed with a short SiO_2 column by eluting with a mixture of hexane and EtOAc (100:10). Identical incubation conditions were used to compare each of the product distribution patterns: substrate 200 μg, pET26b-SHC (5 μg, purified with a Ni^+-NTA column), Triton X-100 (0.2%) in sodium citrate buffer (pH 6.0, 60 mM), total volume 2.0 mL, incubated at 55°C for 6 h. GC conditions: J & D, DB-1, capillary column (length 30 m, I.D. 0.32 mm, film thickness 0.25 μm); injection temperature 300°C; elevated temperature 210─270°C (rate 1°C/min). The ^{1}H-NMR spectra of the peaks (marked with the symbol *) indicated a mixture of materials, and thus, the isolation of some products was unsuccessful due to the marginal production of each product.

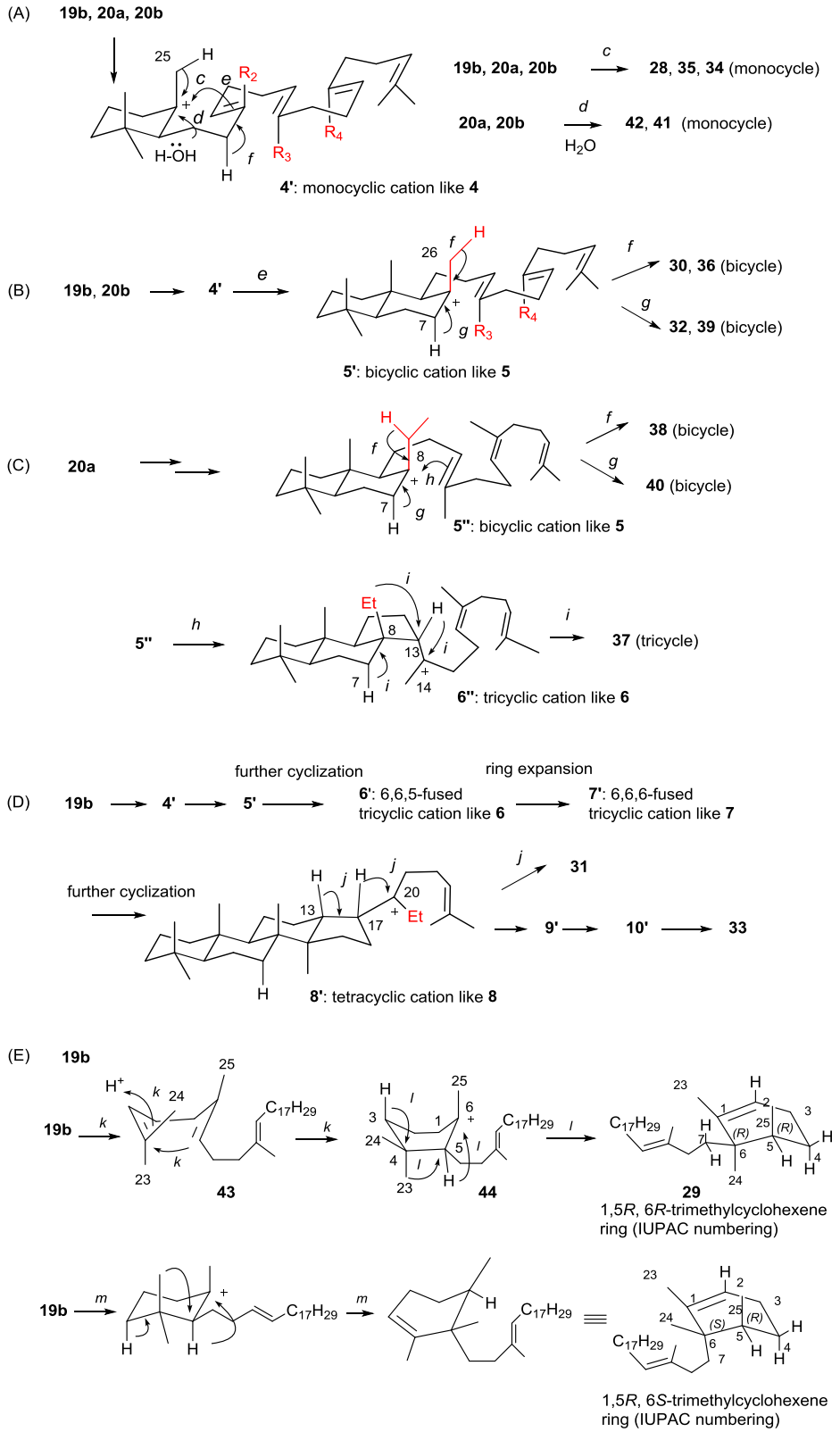
10.1002/ejoc.201800010



34:**35**=1:1.5; **36**:**38**=1:4; **39**:**40**=1:1; **41**:**42**=1:1.4

Fig. 2. Structures of enzymatic products from ethylated analogs **19** and **20**. Et moieties of **19** and **20** were marked with red.

10.1002/ejoc.201800010



Scheme 4. Cyclization mechanisms of **19** and **20,** leading to the productions of **28–42**.

10.1002/ejoc.201800010

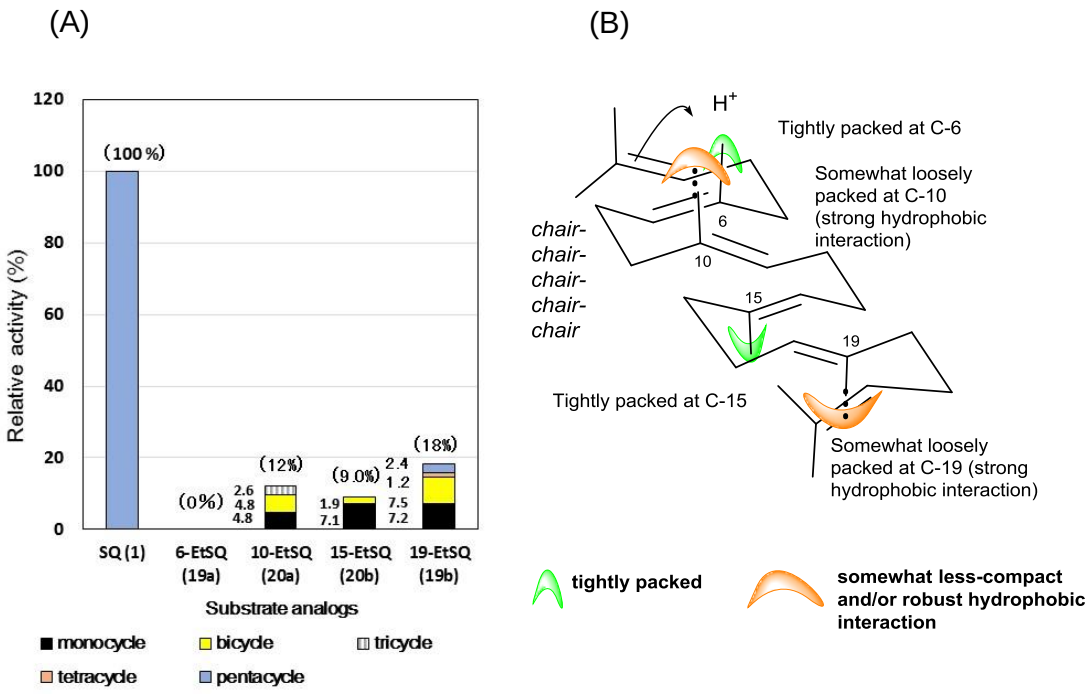
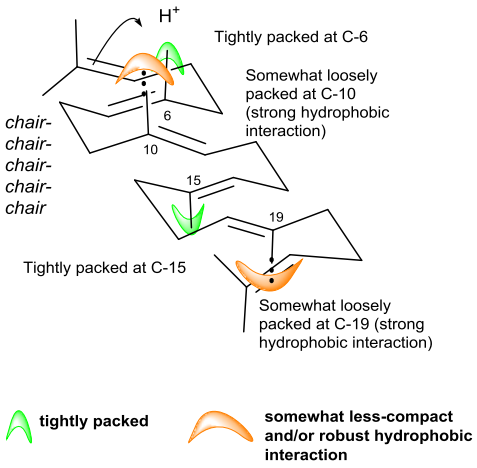


Fig. 3. (A) The activities and product distribution ratios obtained from the incubation experiments of **1**, **19** and **20** (see also Fig. 2). The activities were estimated by incubating each of the substrates at the same conditions as those described in Fig. 2. The total production yields of **1**, **19a**, **19b**, **20a** and **20b** were 100%, 0%, 18%, 12% and 9.0%, respectively. The yield distributions were as follows: monocycle 7.2%, bicycle 7.5%, tetracycle 1.2% and pentacycle 2.4% (total yield 18%) for **19b**; monocycle 4.8%, bicycle 4.8%, tricycle 2.6% (total yield 12%) for **20a**, monocycle 7.1% and bicycle 1.9% (total yield 9.0%) for **20b**.

(B) Recognition (binding) sites at positions C-6, C-10, C-15 and C-19. The bulky Et groups at C-6 and C-15 cannot be accommodated, and thus, the Me-recognition sites are tightly packed. In contrast, the binding sites at positions C-10 and C-19 can partially accommodate the Et group by virtue of the somewhat large cleft size and strong hydrophobic interactions between the cleft and the large Et substituents.

10.1002/ejoc.201800010

TOC (Table of Contents)



Product outcomes of the cyclization of ethyl-substituted squalenes at positions C-6, C-10, C-15 and C-19 by squalene (SQ)-hopene cyclase were investigated. No reaction occurred with 6-EtSQ, and mono- and bicyclic products were accumulated in the reactions with 10- and 15-EtSQ, indicating that the catalytic domain for the A/B/C-ring formation sites is tightly packed. In contrast, in addition to mono- and bicyclic products, 19-EtSQ was converted into tetra- and pentacyclic hopene, indicating that the recognition site for the Me group at C-19 of SQ is somewhat loosely packed and/or a robust hydrophobic interaction occurs between them.