

A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY

CHEM **BIO** CHEM

SYNTHETIC BIOLOGY & BIO-NANOTECHNOLOGY

Accepted Article

Title: Euphorbia tirucalli β -amyrin synthase: critical roles of the steric sizes at Val483 and Met729 and the CH- π interaction between Val483 and Trp534 for the catalytic action

Authors: Tsutomu Hoshino, Kazuya Nakagawa, Yukari Aiba, Daichi Itoh, Chika Nakada, and Yukari Masukawa

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *ChemBioChem* 10.1002/cbic.201700368

Link to VoR: <http://dx.doi.org/10.1002/cbic.201700368>

WILEY-VCH

www.chembiochem.org

A Journal of



***Euphorbia tirucalli* β -amyryn synthase: critical roles of the steric sizes at Val483 and Met729 and the CH- π interaction between Val483 and Trp534 for the catalytic action**

Tsutomu Hoshino,* Kazuya Nakagawa, Yukari Aiba, Daichi Itoh, Chika Nakada, and Yukari Masukawa

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.

E-mail address: hoshitsu@agr.niigata-u.ac.jp

Electronic Supporting Information available: amino acid alignment of various triterpene cyclases, EIMS and NMR spectral data of products **8** and **9**, GC profiles of the W534X mutants, relative enzymatic activities of the site-directed mutants, homology models of the M729X mutants, and primers for preparation of the site-specific enzyme variants.

Accepted Manuscript

Abstract

The functions of Val483, Trp534, and Met729 in *Euphorbia tirucalli* β -amyrin synthase were revealed by comparing the enzyme activities of the site-directed mutants against that of the wild type. The Gly and Ala variants with a smaller bulk size at position 483 dominantly afforded monocyclic camelliol C, suggesting that the orientation of the (3*S*)-2,3-oxidosqualene substrate was not appropriately arranged in the reaction cavity as a result of the decreased bulk size, leading to failure of its normal folding into the chair-chair-chair-boat-boat conformation. The Ile variant with a somewhat larger bulk afforded β -amyrin as the dominant product. Intriguingly, the various variants of Trp534 exhibited significantly decreased enzymatic activities and provided no aberrantly cyclized products, even though the aromatic Phe and Tyr residues were incorporated and the steric sizes of the aliphatic residues were altered. Therefore, the Trp534 residue does not stabilize the transient cation via cation- π interaction. Furthermore, the Trp residue—bearing the largest steric bulk among all the natural amino acids—is essential for high enzymatic activity. A robust CH- π complexation between the Val483 and Trp534 residues is proposed here. Altering the steric bulk at the Met729 position afforded the pentacyclic skeletons. Thus, Met729 is positioned at the E-ring formation site. More detailed insights into the functions of the Val483, Trp534, and Met729 residues are provided by homology modeling.

Key Words: biosynthesis; triterpene; β -amyrin; terpene cyclization; CH- π interaction

Introduction

Polycyclic triterpenes are biosynthesized by (3*S*)-2,3-oxidosqualene cyclases (OSCs).^[1] Their structural diversity is remarkable and about 150 different carbocyclic scaffolds are currently known.^[2] The multi-reaction steps are attained by a single enzyme that generates new C-C bonds and chiral centers in a completely regio- and stereospecific fashion.^[1] β -Amyrin, which is widely distributed in nature, is a pentacyclic triterpene consisting of a 6,6,6,6,6-fused skeleton that is produced by folding the oxidosqualene (**1**) substrate into a chair-chair-chair-boat-boat conformation.^[1i, 3] The biosynthetic pathway of β -amyrin is shown in Scheme 1. In contrast, tetracyclic lanosterol and cycloartenol frameworks are constructed by arranging **1** into a chair-boat-chair-chair conformation.^[4] Squalene-hopene and lanosterol synthases have been investigated in depth,^[1] but progress in the study of β -amyrin synthase has lagged in comparison. We have reported functional analyses of the aromatic active sites in *Euphorbia tirucalli* β -amyrin synthase (*EtAS*),^[5] which are

strictly conserved in all β -amyrin synthases (Fig. S1, Supporting Information). Investigations of triterpene cyclases have indicated that aromatic amino acids mainly stabilize the transient cation produced during the polycyclization reaction.^[6] Our β -amyrin biosynthetic studies, however, have revealed that the conserved aromatic site residues have the following three distinctive functions: (1) stabilization of the cations via cation- π interaction, as revealed by functional analyses of Trp257,^[5c] Tyr259,^[5c] and Phe728^[5a]; (2) maintenance of efficient catalysis by keeping the substrate in the correct position within the reaction cavity, assisted by an appropriate steric bulk, as demonstrated by the analyses of Phe413^[5c] and Phe474,^[5b] and (3) creation of CH- π interaction between the substrate and the aromatic amino side chain, as exemplified by the analysis of Trp257.^[5c,7d,7e] Some important structural units involved in substrate **1** have also been revealed for the normal pathway leading to the production of the β -amyrin scaffold.^[7]

<Scheme 1>

In this paper, we describe the functional analyses of Val483, Trp534, and Met729 in *EtAS*, which are highly conserved in pentacyclic β -amyrin and lupeol synthases (Fig. S1–S3, Supporting Information). We constructed many site-specific mutants by altering the steric sizes of aliphatic residues and/or the type of aromatic residues, and then the relative enzyme activities of the variants against that of the wild-type were evaluated to identify their functions. As shown in Fig. S1, the Val483 in *EtAS* corresponds to Ile or Val in the tetracyclic cycloartenol or lanosterol synthases. The functions of Ile481 in *Arabidopsis thaliana* cycloartenol synthase (*AtCAS1*) and Val454 in *Saccharomyces cerevisiae* lanosterol synthase (*Erg7*) have been reported.^[8a,b] However, the function of the conserved Val residue in pentacyclic β -amyrin synthases has not been reported. The *EtAS* variants V483G and V483A, constructed by us, produced monocyclic camelliol C in a significantly high yield. Trp534 and Met729 are not conserved in the tetracyclic terpene synthases. The homology-modeled structure of *EtAS* suggests that Trp534 and Val483 form a robust CH- π complexation between them to build an organized protein local structure. The Trp534X mutants, irrespective of the aliphatic or the aromatic variants, exhibited significantly decreased enzymatic activity, possibly due to a collapse of the organized protein local structure. The role of Met729 residue in the OSC-mediated polycyclization reaction is unknown. A decrease in the steric bulk at position 729 gave aberrant cyclization products. Nevertheless, the variant having almost the same steric size as Met, did not yield the aberrant cyclization products. Therefore, the appropriate steric volume at the three positions is crucial in

correctly positioning the substrate in the reaction cavity. Herein, we report the detailed functional analyses of the three residues. Homologically modeled variant structures at the three sites provide more detailed insights into their mechanistic functions.

Results and discussion

Identification of triterpene products 8–15

Four site-directed variants were constructed by substituting Val483 with Gly, Ala, Ile, and Phe. Seven variants were constructed for the Trp534 position by replacing it with the substituents Ala, Val, Ile, Met, His, Phe, and Tyr. Seven mutants were produced using Gly, Ala, Val, Leu, Phe, Trp, and Asn in lieu of Met729. Mutagenesis of these positions in the wild-type pYES2-*EtAS*/CT, constructed earlier by our group,^[9] was performed using the QuikChange site-directed mutagenesis method. These mutants were expressed in 100 mL cultures of *Saccharomyces cerevisiae* GIL77, and the triterpene products accumulated in the yeast cells were extracted with hexane after saponification with 15% KOH/MeOH. The hexane extracts were dried under the reduced pressure and subjected to gas chromatography (GC) analyses (Figs. 1A and 1B). Compounds **10** and **11** were produced by the wild-type *EtAS* and previously identified as butyrospermol and tirucalla-7,24-dien-3 β -ol, respectively.^[9]

As shown in Fig. 1A (V483X variants), compounds **8** and **9**, which are not produced by the wild type, were detected from the V483G and the V483A mutants. A large-scale culture (24 L) was conducted to identify the structures of **8** and **9**. The cultured cells were collected by centrifugation, and the triterpenes were extracted with hexane after saponification. The extracts were loaded onto a SiO₂ column for chromatography using a mixed solvent of hexane:EtOAc (= 100:15), yielding pure compound **8**. Compound **9** was produced in a very small amount, rendering its isolation difficult. Repeated 5% AgNO₃-SiO₂ column chromatography (hexane:EtOAc = 100:15 and 100:7.5) finally gave the **9**-enriched fraction. This fraction was acetylated with Ac₂O/Py, and the acetate fraction was subjected to normal-phase high-performance liquid chromatography (hexane:tetrahydrofuran = 100:0.5), affording **9**-acetate in an almost pure state. The two products were subjected to nuclear magnetic resonance (NMR) analyses, including two-dimensional NMR (Figs. S2 and S3). In the ¹H-NMR spectrum of **8** (600 MHz, CDCl₃; Fig. S2.2), six olefinic methyl (Me) protons and five olefinic protons were observed, indicating that product **8** possesses the same number of double bonds as that of substrate **1**. However, an alcoholic carbon was found at δ_c 75.0 ppm (d, C-3) in the ¹³C-NMR spectrum (150 MHz). Thus, product **8** is suggested to be a

monocyclic compound. In the heteronuclear multiple-bond correlation (HMBC) spectrum, Me-23 (δ_{H} 1.04, 3H, s) and Me-24 (δ_{H} 0.901, 3H, s) had cross-peaks with C-3 and C-5 (δ_{C} 48.9, d). Furthermore, olefinic Me-25 (δ_{H} 1.79, 3H, s) had strong HMBC correlations with C-5 (δ_{C} 48.9, d) and C-6 (δ_{C} 137.1, s), and with C-1 (δ_{C} 118.3, d). These data strongly support a cyclohexene ring in the molecular structure of product **8**. Detailed analyses of the COSY, HOHAHA, NOESY, HSQC, and HMBC spectra (Figs. S2.4–S2.8) together with the distortionless enhancement by polarization transfer (DEPT) spectra definitively revealed compound **8** to be camelliol C. The assignments of the NMR data are shown in Fig. S2.9. In the ^1H -NMR spectrum (400 MHz, CDCl_3) of product **9**-acetate (Fig. S3.2), five olefinic Me protons and four olefinic protons were found, in addition to methylenic protons (δ_{H} 4.99, s; 4.82, s, H-25, each 1H). The presence of the methylenic group was further confirmed by observing the signal at δ_{C} 109.4 ppm (t) in the ^{13}C -NMR (Fig. S3.3) and DEPT 135 spectra (100 MHz, C_6D_6). In the HMBC spectrum (Fig. S3.8), H-25 had definitive cross-peaks with C-5 (δ_{C} 51.7, d), C-6 (δ_{C} 147.1, s), and C-1 (δ_{C} 31.9, t). Me-23 (δ_{H} 1.05, 3H, s) and Me-24 (δ_{H} 0.985, 3H, s) had clear HMBC cross-peaks with C-5, C-3 (δ_{C} 78.3, d), and C-4 (δ_{C} 39.4, s). These analyses indicate that product **9** is monocyclic achilleol A (Fig. S3.9).

Gas chromatograms of the W534X variants are shown in Fig. S4. No aberrant cyclization products were found. The product patterns were the same as that of the wild type; namely, β -amyrin **2**, butyrospermol **10**, and tirucalla-7,24-diene-3 β -ol **11**. The product distribution ratios of the W534X mutants are shown in Table S1. As shown in Fig. S4, production of **2** by all the enzyme variants, including the aliphatic and aromatic mutants, was significantly decreased compared with that produced by the wild type; thus, minor products **10** and **11**, which are generated by the wild type, were scarcely detected in the GC profiles of certain mutants (Fig. S4 and Table S1).

Fig. 1B depicts the GC analyses of triterpene fractions from various M729X variants. Products **12**, **13**, and **15** were identified as lupeol, germanicol, and ψ -taraxasterol, respectively,^[5] by comparing their GC/mass spectrometry (MS) spectra (MS fragment pattern and retention times) with those of our previously identified authentic samples.^[5] The MS fragment patterns for the acetylated products of the M729G variants are shown in Fig. S5.

Cyclization pathway from substrate **1** into products **8**–**15**

The product structures and the cyclization pathway from **1** are illustrated in Scheme 2. Substrate **1** was cyclized by the V483G and V483A mutants to yield the monocyclic cation **3**. Deprotonation of H-1 and Me-25 afforded camelliol C **8** and achilleol A **9**,

respectively. Further cyclization reaction via a chair-chair-chair-*boat* folding conformation gave the tetracyclic dammarenyl cation **4**. By contrast, the cyclization reactions through a chair-chair-chair-*chair* conformation provided the 17-*epi*-dammarenyl cation **4'**. These tetracyclic cations **4** and **4'** underwent backbone rearrangements (H-17→C-20, H-13β→C-17, Me-30α→C-13, and Me-18β→C-14), followed by proton elimination of H-7α to yield **10** and **11**. A ring expansion of **4** afforded the baccharenyl cation **5**, followed by formation of the lupanyl cation **6**.

The mutations at Met729 gave rise to an aberrant cyclization pathway. A decrease in the steric bulk (the Gly, Ala, and Val variants) halted the polycyclization cascade at the lupanyl cation **6** stage; the deprotonation of Me-29 or Me-30 furnished lupeol **12**. The lupanyl cation **6** underwent a ring expansion to yield the final oleanyl cation **7**, but germanicol **13** and ψ-taraxasterol **15** were produced in addition to the normal cyclization product **2**. H-18α of **7** was subjected to proton elimination to give **13**. 1,2-Shift of Me-29 to C-19 gave the taraxasteryl cation **14**, followed by abstraction of H-21β to yield **15**.

<Scheme 2>

Functional analyses of the Val483 residue

Table 1 shows the product distribution ratio obtained by mutating the Val483 residue. Decrease in the steric bulk (Gly and Ala variants) dramatically altered product distribution and the aberrant cyclization products. A significantly large quantity of monocyclic **8** and a very small amount of monocyclic **9** were produced along with the correctly cyclized pentacyclic **2** (**2**:**8**:**9** = 1.7:18.1:1 for the V483G mutant). However, the Ile variant did not yield the two aberrant products; instead, the amount of **2** produced by this mutant was nearly the same as that produced by the native enzyme (Table 1). It is noteworthy that **10** and **11** are produced by the wild-type *EtAS*, as we reported before,^[9] and the production yields by the V483X mutants remained almost unchanged (Table 1), indicating that the alteration in steric bulk at the Val483 position had little influence on formation of the tetracyclic cation **4**. Fig. 2 compares the relative enzyme activities of the V483X variants with that of the wild-type *EtAS*. The total amounts of the triterpene products, which were estimated by GC analyses, were divided by the enzyme amounts expressed *in vivo*, which were determined by western blot analyses (Figs. S6.1–S6.3). This procedure has been successfully applied to the estimation of the relative enzymatic activities of the mutants against that of the native *EtAS*.^[1i,5,9] The relative activities for the total amounts of triterpenes produced increased

in proportion to the increase in steric bulk (van der Waals volume): 65% for the Gly mutant, 82% for the Ala mutant, and 89% for the Ile mutant. The relative activities for the production of β -amyrin also increased with larger steric bulk (see the right-pointing arrow in Fig. 2). In contrast, the production of monocyclic **8** and **9** was inversely proportional to the increase in the van der Waals volume;^[10] no production of **8** and **9** by the Ile mutant was found, although the steric size of Ile is somewhat larger than that of Val (the wild type). The decreased steric bulk at position 483 could possibly result in an inaccurate arrangement of the substrate near the B-ring formation site within the reaction cavity, leading to polycyclization abortion at the monocyclic stage and the decreased enzyme activity for the total amounts of triterpenes produced. The aromatic Phe mutant had no enzymatic activity (Fig. 2), which would have occurred as a result of the failure to achieve CH- π complexation, leading to a collapse of the organized protein local structure, as discussed below in the section on the homology-modeled structures. Intriguingly, production of **8** and **9** is also found by mutating the corresponding amino acids of Erg7 and AtCAS1. In the case of Erg7, site-specific mutagenesis of Val454 into Gly and Ala gave **9** only, besides lanosterol.^[8b] On the other hand, the mutation of Ile481 in AtCAS1 into Gly and Ala afforded both **8** and **9**, but the production yield of **9** was significantly higher than that of **8**.^[8a] Joubert et al. suggested that the steric size of Val (or Ile) residue may participate in prefolding the C-10–C-11 double bond of **1** to facilitate the formation of chair-shaped B-ring.^[8b] In contrast, an overwhelming production of **8**, compared with that of **9**, was observed for the *EtAS* V483X variants. At the present time, we have no explanation for this difference of product distribution among the different triterpene cyclases. The Matsuda group discovered camelliol C synthase (*AtCAMS1*) from *A. thaliana*, which produces **8** in an overwhelming yield (98%), besides **9** (2%) and **2** (0.2%).^[11] Amino acid alignment study indicated that the corresponding position is substituted with Ala,^[11] despite Val and Ile being conserved in tetracyclic and pentacyclic triterpene cyclases (see Figs. S1.1 and S1.3). The Matsuda group also reported that *AtCAMS1* is a descendant of the enzymes that form pentacyclic triterpenoids.^[11] Mutations of this site in *EtAS* corroborate the proposal of the Matsuda group that camelliol C synthase evolved from β -amyrin synthases. Furthermore, the triad consisting of the Val483–Trp534–Tyr259 motif (*EtAS*), of which Trp and Tyr residues are conserved in *AtCAMS1*, as shown in Fig. S1.3. This fact may provide additional evidence that camelliol C synthase evolved from 6,6,6,6,6- or 6,6,6,6,5-fused pentacyclic triterpene synthase.

<Table 1 and Fig. 2>

Functional analyses of the Trp534 residue

Figs. 3 and S7.2 show the relative enzymatic activities of the aliphatic and aromatic variants targeted for residue Trp534. All the aliphatic variants showed significantly decreased activities (0.8–5.0%). Enhancement of the activities in the Phe and Tyr variants was slight or negligible, and no significant difference between the aromatic and aliphatic variants (e.g., the Met and Ile mutants) was observed (see also Fig. S7.2). By contrast, the *EtAS* variants of Trp612 into Phe or Tyr exhibited the relative activities of 45–65% against that of the wild-type, but little activities for the aliphatic mutants, in addition to the generation of aberrant cyclization products (unpublished results by us), which are significantly different from the experimental results of Trp534 variants (Figs. 3 and S7.2). The homology modeling shows that the Trp612 residue is *horizontal* to the lanosterol ligand.^[11] Thus, the π -electrons of Trp612 can stabilize the cationic intermediate via cation- π interaction. This distinct behavior between Trp534 and Trp612 mutants supports that Trp534 residue is *perpendicular* to the lanosterol ligand as shown in the homology modeling (depicted below). As described in the section of the homology modeling, the Val483 residue is supposed to participate in the CH- π interaction with the Trp534 that is arranged perpendicularly to the ligand. It is likely that this complexation build the rigidly organized protein architecture around the A/B/C-ring formation sites in the enzyme cavity. The CH- π interaction between the Val and the Trp is most robust compared with that between Val and Phe or Val and Tyr, due to the higher π -electron density of Trp relative to that of Phe and Tyr. Thus, a robustly organized protein architecture would be created. No aberrant cyclization products, such as **8**, **9**, **12**, **13**, or **15**, were generated by all the mutants examined (Fig. S4 and Table S1). The significant disruption of the robust protein architecture around the B/C-ring formation sites, which results from the mutations at this position, would have led to the significantly decreased enzyme activity and absence of aberrant cyclization product generation. In addition to the robust CH- π complexation of Trp534 with Val483, the Trp534 residue possessing the largest bulk size among natural amino acids may form a tightly packed cavity to enable close contact with substrate **1** around A/B/C-ring formation sites, by which the desired folding conformation is acquired to facilitate the polycyclization reaction. The steric size at this position of the Phe and the Tyr variants are smaller than that of the wild-type (Trp residue), which could prevent the folding of the desired conformation and yield the significantly decreased activities of the Phe and Tyr variants, that is, the steric size at position 534 also could be crucial for the normal catalytic action. Thus, the integral structural unit, formed through the CH- π

complexation between the Val and the Trp, builds a tight packing cavity to have close contact with substrate **1** around A/B/C-ring formation sites, as discussed in the section of the homology model.

<Fig. 3>

Functional analyses of the Met729 residue

As illustrated in Scheme 2, alteration of the steric bulk at this position renders germanicol **13** and ψ -taraxasterol **15**, which are generated from the oleanyl cation **7**. In addition, lupeol **12** was produced from the lupanyl cation **6**. In other words, this Met residue is located at the E-ring formation site, and its appropriate steric bulk enables the accurate placement of the substrate in the reaction cavity, leading to its correct folding (a boat conformation) around the site. Indeed, the production of β -amyrin was roughly increased in proportion to the increase in van der Waals volume (see the right-pointing arrow in Fig. 4). Furthermore, the increments of the total amounts of triterpenes produced were in proportion to the steric size, as shown in Fig. 4; namely, 69.9% for the Gly mutant and 101.1% for the Leu variant possessing nearly the same steric bulk as the native *EtAS*. However, the activity of the Phe mutant decreased to 80.2% due to its size being larger than that of the wild type. The Trp mutant, bearing too large a side chain, exhibited no enzyme activity. In our experiments, Phe728, positioned prior to Met729, stabilized the oleanyl cation **7**.^[5a] Decreasing of the steric bulk at this position afforded lupane **12** (5-membered E-ring), germane **13** and ursane **15** scaffolds (different deprotonation products), suggesting that the steric size of Met may have participated in impelling a correct boat conformation around the E-ring formation site to yield the 6-membered E-ring (oleanyl cation **7**) from the 5-membered E-ring (premature lupanyl cation **6**). The distorted boat structure given by the decreased steric bulk would have generated the different deprotonation products according to Scheme 2. The expressed quantities of the mutated *EtAS*s and of each triterpene produced are shown in Figs. S8.1–S8.7. In the tetracyclic cycloartenol and lanosterol synthases, the Asn residue is highly conserved at this position (see Figs. S1.1 and S1.2). The Asn mutant produced **13** in a significantly high distribution ratio (54%), but the production of tetracyclic products **10** and **11** did not increase. Thus, the substitution of Met with Asn does not contribute at all to the production of the tetracyclic triterpenes.

<Table 2 and Fig. 4>

Mechanistic insights from the homology-modeled structures

No X-ray structure of β -amyrin synthase has been reported hitherto. To gain a deeper insight into the functions of the Val483, Trp534, and Met729 residues, a homology models, based on the X-ray structure of human lanosterol synthase (*HsLAS*),^[13] were constructed, in which the three aforementioned residues were highlighted in Fig. 5A. The Me groups of Val483 are perpendicular to the indole ring, so a strong CH- π interaction may occur between V483 and Trp534, and organized local protein structures are assembled around the A/B/C ring formation sites. Moreover, Val483 is situated near the lanosterol ligand (ca. 3–4 Å), suggesting that this residue plays a vital role in the substrate **1** polycyclization reaction. A red arrow indicates the position of Trp534 in the cavity. The indole ring of Trp is perpendicular to the ligands. Brandl et al. reported that CH- π interaction is observed in various protein structures when the distance between the carbon atom and the mass center of the aromatic system (d_{C-X}) is ca 3.4–4.4 Å.^[11a] As shown in Fig. 5A, the d_{C-X} are 3.9–4.3 Å for the wild-type EtAS. Furthermore, the shortest contact distance (3.5–3.9 Å) is also reported as examples of CH- π interaction in proteins.^[11b, c] In the case of EtAS, 3.6 Å is estimated as the shortest distance between the Me group and the indole ring (a green solid line in Fig. 5A). Therefore, CH- π complex between V483 and Trp534 residues can be credible. Met729 is also located near the ligand around the E-ring formation site (ca. 3–4 Å). Fig. 5B illustrates the arrangements of the three active site residues on substrate **1**, which is folded in a normal chair-chair-chair-boat-boat conformation. Figs. 6 and 7 show the structures of V483X and W534X, respectively. Models of the M729X variants are shown in Fig. S9. The orientations of the mutated residues are optimized in the models by minimizing the steric hindrance between the ambient- and mutated residues.

As depicted in Fig. 6, the CH- π interaction in the Gly mutant (A) is minimal, because the Me group is missing, and a larger space (marked with a blue oval) is generated between the substrate and the Gly residue. This vacant space and the distortion of the organized local structure, which could result from the mobility of the Trp534, could lead to less contact with substrate at B-ring formation site and halt the polycyclization reaction at premature monocyclic stage (production of monocycles **8** and **9**). For the Ala mutant (C), the Ala residue has the Me group and thus CH- π complexation is possible, but this interaction is weaker than that created by the Val residue (wild-type) because only one Me group is involved. Furthermore, the smaller steric size could afford the free space around the A/B-ring sites, and thus the premature cyclization products **8** and **9** were also produced in a similar way as by the Gly mutant. For the Ile mutant (D), the free space is minimal between the Ile residue and the

substrate, and the local protein architecture is maintained because of the strong CH- π interaction between the two Me groups of Ile and the indole ring. In the case of the Phe variants (E and F), the embedded Phe residue with its large steric bulk pushes the Trp534 residue (labeled as W534 in Fig. 6) aside (due to the collision between Phe and Trp534) and CH- π complexation therefore fails, leading to disorganization of the local structure. The models give additional evidence for the crucial role of the steric size in the normal polycyclization cascade, which is consistent with the outcomes illustrated in Fig. 2.

As depicted in Fig. 7, the Trp534 residue is arranged in proximity to the B/C-ring formation sites and perpendicularly to the ligand (Fig. 5). Thus, the function of the Trp534 residue is not to offer the π -electrons to the intermediary cation produced during the polycyclization reaction (cation- π interaction), but to place the substrate at the correct position in the reaction cavity (by virtue of the residue's appropriate steric size) in order to furnish a chair-chair-chair conformation for the fused A/B/C-ring system. The aliphatic mutants replaced by Ala, Ile, and Met are not capable of forming the CH- π complexation, because π -electrons are lacking at position 534, and thus the organized local structure at the A/B/C-ring sites collapses. Furthermore, large empty spaces (depicted as blue-filled ovals in Fig. 7) are generated, as shown in the models of the aliphatic mutants (B, C, and D). This indicates that the substrate is moving freely within the reaction cavity, resulting in failure to organize its normal folding conformation. The models for the Phe and Tyr mutants (aromatic variants) are depicted as (E) and (F), respectively, in Fig. 7. Because these aromatic substituents are not arranged perpendicularly to the ligand (as shown by the green arrows), they also fail to bind the substrate tightly in the cavity, leading to a significantly decreased enzyme activity (<10% of the wild-type activity; see Fig. 3).

The models of the mutated *EtASs* targeted for the Met729 position are shown in Fig. S9. The Gly and Ala mutants have a significantly large free space (blue-filled ovals) between the residues and the E-ring site. The Val mutant has a somewhat compact space. The Leu variant has a similar space as the wild type (tightly packed), because the side chains of Leu and Met possess a similar van der Waals volume (see Fig. 4).^[10] The Phe and Asn residues are oriented relatively far from the E-ring formation site (see the green arrows in Fig. S9; the long axes of the side chains differ from that of Met). As a result, these mutations have decreased enzyme activities and produce aberrant cyclization products (**12**, **13**, and **15**). The Asn mutant has a larger space than the Phe variant and may generate greater amounts of the aberrantly deprotonated product **13**. The homology-modeled structures further support the theory

that the steric bulk at site 729 is essential in the normal polycyclization pathway, as discussed above (see Fig. 4).

The Val483 and Trp534 residues are exclusively conserved in the β -amyryn and lupeol synthases, which produce the pentacyclic skeletons (Figs. S1.1 and S1.3). The Val or Ile residue is also highly conserved in the cycloartenol and lanosterol synthases at the position equivalent to Val483 in *EtAS* (Figs. S1.1 and S1.3). Trp534 in *EtAS* corresponds to Tyr in the cycloartenol and lanosterol synthases. The Wu group has reported the importance of His234 and Tyr510 in Erg7, and further pointed out that a hydrogen-bonded complex is formed by these two residues.^[1f, 13] We propose that this Tyr residue forms the CH- π complex with the Val or Ile residues, in addition to the hydrogen-bond complex with His (Fig. S10, ESI); these robust interactions build an organized local structure through the A–D-ring formation sites. It was proposed that the function of the hydroxyl group of Tyr is to abstract H-9 of the final protosteryl cation for lanosterol formation.^[14] However, because the Tyr \rightarrow Phe mutation still afforded lanosterol, despite the loss of the deprotonating OH group,^[15] the function of the Tyr remains unknown. Fig. S10 shows the structure of human lanosterol synthase (*HsLAS*) superimposed onto that of *EtAS*. The three residues in *HsLAS* are in close proximity to each other (ca. 3-4 Å) but the side chains of Tyr503 and Trp534 are oriented very differently (green and orange arrows in Fig. S10). The disposition of Tyr503 is essentially the same as that indicated in Fig. 7F. The CH- π interaction between Val and Tyr is weaker than that between Val and Trp because the former has a lower π -electron density than the latter. Nevertheless, the local structure of the tetracyclic terpene cyclases is reinforced by hydrogen bonding with His232 (shown in green in Fig. S10). Thus, a more robust CH- π interaction occurs in *EtAS* (Fig. S10, ESI) than in *HsLAS*. Functional analyses of Val483 and Trp534 in *EtAS* indicate that the CH- π interaction between these residues accounts for the robustly organized reaction cavity around the A/B/C-ring formation sites. This configuration is necessary for *EtAS*-mediated cyclization. By contrast, the Val (Ile)–Tyr–His triad complex forms a robust protein architecture around the A–D-ring formation sites in lanosterol and cycloartenol biosynthesis. Fig. S1.3 show the multiple alignment of amino acid sequences for various OSCs. Protostadienol synthase also folds substrate **1** into a chair-boat-chair-chair conformation. In contrast, the dammarenediol-, baccharis oxide-, baruol-, and thalianol synthases force **1** into a chair-chair-chair folding conformation. As discussed in Fig. S11, we propose that the Val453–Tyr503–His-232 triad complex (*HsLAS* numbering) may direct **1** into a chair-boat-chair-chair conformation, notably to shape a boat

structure around the B-ring formation site. By contrast, the Val483–Trp534–Tyr259 triad (*EtAS* numbering) would probably impel **1** into the chair-*chair*-chair folding conformation, notably to shape a chair structure around the B-ring formation site. Mutations of Val and Ile into Gly and Ala in lanosterol and cycloartenol synthases also afford monocycles **8** and **9**.^[8a, b] This would have occurred by the disruption of the CH- π interaction between the Val (Ile) and the Tyr residues (in a similar way as the *EtAS* homology modelling shown in Figs. 6B and 6C and Fig. S10). As described above, the different orientations and the steric volumes of the side chains of Tyr503 and Trp534 (see Figs. 5 and S10) may confer the different folding conformation (either pre-boat or pre-chair structure). However, no experimental evidence for this idea is provided at the present time. To validate this theory and to clarify the molecular mechanism for the boat- or chair- conformational selectivity, further studies are required.

Conclusions

The present study highlights the importance of the steric bulk of the aliphatic amino acids at positions 483 and 729 for normal folding of the substrate to complete the polycyclization cascade, and for the deprotonation position in order to lead to the β -amyrin scaffold as a normal end-product. We propose that Val483 and Trp534 form a robust CH- π complex between the active site residues. This is the first study to report that the CH- π interaction has a crucial role in reinforcing the protein architecture of OSCs in order to ensure the correct placement of substrate **1** within the active site in the reaction cavity. Furthermore, it was noted that all the site-specific variants of position Trp534 had little or no enzymatic activity and afforded no aberrant cyclization products, which differed significantly from the behaviors of the V483X and M729X variants. This finding reinforces the fact that the Trp534 residue has a crucial role in building the rigidly organized protein architecture around the A/B/C-ring formation sites, which occurs through its CH- π complexation with the Val483 residue. The insights derived from the homology-modeled structures further are consistent with the functions inferred from the various mutants of Val483, Trp534, and Met729.

Acknowledgments

This work was supported by JSPS KAKENHI (Grant Nos. 25450150 and 18380001).

Experimental Section

Instruments

NMR spectra were recorded in C_6D_6 with a Bruker DMX 600 or DPX400 spectrometer. The chemical shifts (δ) are given in ppm relative to the residual solvent peak as the internal reference ($\delta_H=7.28$ and $\delta_C=128.0$ ppm for C_6D_6 , and $\delta_H=7.26$ and $\delta_C=77.0$ ppm for $CDCl_3$). GC analyses were done on a Shimadzu GC-2014 chromatograph fitted with a flame ionization detector (J&W DB-1 capillary column, 0.25 mm x 30 m). GC/MS spectra were obtained with a JMS-Q1000 GC K9 (JEOL) instrument under electronic impact at 70 eV with a Zebron ZB-5ms capillary column (0.25 mm x 30 m), the oven temperature being elevated from 220 to 270 °C (3 °C/min).

Mutagenesis experiments

Mutagenesis of V483, W534 and Met729 in wild-type pYES2-*EtAS*/CT was performed with the QuikChange site-directed mutagenesis method, according to the previous report.⁹ The oligonucleotide primers used were shown in Table S2. The integrity of the mutated gene was verified by sequencing.

Expression of the mutated *EtAS* in *S. cerevisiae* GIL77, and analyses of triterpene amounts produced.

The expression plasmid pYES2-*EtAS* was transformed into *S. cerevisiae* GIL77 strain lacking lanosterol synthase (*erg7*, *ura3-167*, *hem3-6*, *gal2*), by using Frozen-EZ Transformation II Kit (Zymo Research). The transformants were plated onto synthetic complete medium without uracil (SC-U), supplemented with ergosterol (20 μ g/mL), hemin chloride (13 μ g/mL), and Tween 80 (5 mg/mL), and then incubated at 30°C for 3 days. Expression of *EtAS* in the yeast was conducted as reported by Kushiuro et al.^[16] A 100 mL culture of *S. cerevisiae* GIL77 harboring pYES2-*EtAS* was grown in SC-U medium that was supplemented with ergosterol (20 μ g/mL), hemin chloride (13 μ g/mL), and Tween 80 (5 mg/mL) at 30°C with shaking (150 rpm). After grown for 2 days, cells were collected and resuspended in SC-U medium without glucose that was supplemented with ergosterol, hemin chloride, and Tween 80, and then 2% galactose was added to induce *EtAS* expression at 30 °C for 24 h. The yeast cells were collected and resuspended in 0.1 M potassium phosphate (pH 7.0) supplemented with 3% glucose and hemin chloride, and further incubated at 30 °C for 24 h. To the grown cells collected, was added 15% KOH/MeOH and the suspension was refluxed by heating to disrupt the cells. The lipophilic materials including the cyclization products were extracted with hexane, the extract having being subjected to GC/MS and GC analysis for identifying the triterpenes produced and for quantifying the triterpene produced,

respectively. As the internal reference, geranylgeraniol was used for the quantification of the triterpene products.

Evaluation of the enzyme activities for the wild- and mutated *EtASs*.

The quantitative assays of the expressed *EtASs* were performed by Western blot analyses according to the previous paper.^[9] To minimize the experimental errors, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used. GAPDH is widely known to be housekeeping protein and frequently used as the internal reference.^[17]

Spectroscopic data of products 8 and 9.

Product 8: Oil; ¹H-NMR (600 MHz, CDCl₃): δ=0.900 (s, 3H, Me-24), 1.04 (s, 3H, Me-23), 1.42 (m, 1H, H-7), 1.67 (s, 9H, Me-27, Me-28 and Me-29), 1.71 (s, 3H, Me-26), 1.75 (s, 3H, Me-30), 1.79 (4H: s, 3H for Me-25; m, 1H for H-7), 2.06 (m, 6H: 1H for H-2, 4H for H-15 and H-19, and 1H for H-8), 2.10 (m, 4H, H-11 and H-12), 2.14 (m, 4H for H-16 and H-20), 2.24 (m, 1H, H-8), 2.31 (m, 1H, H-2), 3.53 (t-like, *J*=7.0 Hz, H-3), 5.24–5.15 (4H, m, H-10, H-13, H-14 and H-21), 5.31 (br s, 1H, H-1); ¹³C-NMR (150 MHz, CDCl₃): δ=15.98 (q), 16.04 (q) and 16.09 (q, 2C) that are assignable for Me-24, Me-26, Me-27 and Me-28, 17.7 (q, Me-29), 22.6 (q, Me-25), 25.3 (q, Me-23), 25.7 (q, Me-30), 26.7 (t) and 26.8 (t) for C-16 and C-20, 27.2 (t, C-7), 28.2 (t, 2C) for C-11 and C-12, 31.8 (t, C-2), 38.1 (s, C-4), 39.7 (t, 2C, C-15 and C-19), 42.0 (t, C-8), 48.9 (d, C-5), 75.0 (d, C-3); 118.3 (d, C-1), 122.42 (d), 124.24 (d), 124.4 (d) and 124.7 (d) that are assignable for one carbon among C-10, C-13, C-17 and C-21; 131.2 (s, C-22); 134.9 (s), 135.2 (s), 135.4 (s) that are assignable for one carbon among C-9, C-14 and C-18; 137.1 (s, C-6); MS (EI): *m/z* (%): 69 (76), 81 (100), 95 (36), 109 (35), 121 (55), 136 (34), 149 (17), 203 (13), 259 (1.3), 274 (2.4), 339 (0.8), 357 (1.2), 408 (1.2), 426 (M⁺, 1.4).

Product 9 acetate: Oil; ¹H-NMR (400 MHz, C₆D₆): δ=0.985 (s, 3H; Me-24), 1.05 (s, Me-23), 1.69 (s, 3H, Me-29), 1.73 (s, Me-27 or Me-28), 1.75 (s, 3H, Me-27 or Me-28), 1.69 (m, 1H, H-2), 1.81 (bs, 3H, Me-30), 1.81 (3H, 1H for H-5 and 2H for H-7, buried in the large Me-30 signal), 1.82 (s, 3H, COMe), 1.96 (m, 1H, H-2), 2.00 (m, 1H, H-1), 2.02 (m, 1H, H-8), 2.23 (m, 4H, H-15 and H-19), 2.28 (m, 4H, two CH₂ protons among H-11, H-12, H-16 or H-20), 2.28 (m, 4H, two CH₂ protons among H-11, H-12, H-16 or H-20), 2.31 (m, 1H, H-8), 2.33 (m, 4H, two CH₂ protons among H-11, H-12, H-16 and H-20), 4.82 (s, 1H, H-25), 4.96 (m, 1H, H-3), 4.97 (s, 1H, H-25), 5.37 (1H, t, *J*=6.8 Hz, H-21), 5.42 (m, 3H, H-10, H-13 and H-17); ¹³C-NMR (100 MHz, C₆D₆): δ=16.1 (q, one carbon among Me-26, Me-27 and Me-28), 16.2 (2C, q, two carbons among Me-26, Me-

27 and Me-28), 17.3 (q, Me-24), 20.7 (COCH₃), 24.4 (t, C-7), 25.8 (Me-30), 26.2 (q, Me-23), 27.1 (1C, t, either C-11, C-12, C-16 or C-20), 27.3 (1C, t, one carbon among C-11, C-12, C-16 and C-20), 28.7 (1C, t, one carbon among C-11, C-12, C-16 and C-20), 28.8 (1C, t, one carbon among C-11, C-12, C-16 and C-20), 29.1 (t, C-2), 31.9 (t, C-1), 38.9 (t, C-8), 39.4 (s, C-4), 40.2 (2C, C-15 and C-19), 51.7 (d, C-5), 78.3 (d, C-3), 109.4 (t, C-25), 124.8 (d, 2C), 124.9 (2C, d), 131.1 (s, C-22), 134.9 (s), 135.3 (s), 135.5 (s), 147.2 (s, C-6), 169.6 (COCH₃). MS (EI): *m/z* (%): 69 (100), 81 (82), 95 (28), 107 (23), 121 (21), 136 (18), 175 (20), 203 (14), 257 (1.1), 271 (2.1), 339 (2.3), 365 (1.1), 408 (2.2), 468 (M⁺, 1.0).

References

- [1] Reviews for the polycyclization of triterpene cyclases. (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) T. Hoshino, *Org. Biomol. Chem.*, **2017**, *15*, 2869–2891.
- [2] R. Xu, G. C. Fazio and S. P. T. Matsuda, *Phytochemistry* **2004**, *65*, 261–291.
- [3] (a) A. Eschenmoser, L. Ruzicka, O. Jeger and D. Arigoni, *Helv. Chim. Acta.* **1955**, *38*, 1890–1940. (b) T. Suga and T. Shishibori, *Phytochemistry*, **1975**, *14*, 2411–2417.
- [4] (a) A. Eschenmoser and D. Arigoni, *Helv. Chim. Acta*, **2005**, *88*, 3011–3050. (b) E. J. Corey and S. C. Virgil, *J. Am. Chem. Soc.*, 1991, **113**, 4025–4026. (c) T. Hoshino, A. Chiba and N. Abe, *Chem. Eur. J.*, **2012**, *18*, 13108–13116.
- [5] (a) R. Ito, I. Hashimoto, Y. Masukawa, *Chem.-Eur. J.*, **2013**, *19*, 17150–17158. (b) R. Ito, Y. Masukawa, C. Nakada, K. Amari, C. Nakano and T. Hoshino, *Org. Biomol. Chem.*, **2014**, *12*, 3836–3846. (c) R. Ito, C. Nakada and T. Hoshino, *Org. Biomol. Chem.* **2017**, *15*, 177–188.
- [6] 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, and references cited therein.
- [7] (a) I. Abe, Y. Sakano, H. Tanaka, W. Lou, H. Noguchi, M. Shibuya and Y. Ebizuka, *J. Am. Chem. Soc.*, **2004**, *126*, 3426–3427. (b) I. Abe, Y. Sakano, M. Sodeyama, H. Tanaka, H. Noguchi, M. Shibuya and Y. Ebizuka, *J. Am. Chem. Soc.*, **2004**, *126*, 6880–6881. (c) T. Hoshino, Y. Yamaguchi, K. Takahashi and R. Ito, *Org. Lett.*, **2014**, *16*, 3548–3557. (d) T. Hoshino, Y. Miyahara, M. Hanaoka, K. Takahashi and I. Kaneko, *Chem. Eur. J.*, **2015**, *21*, 15769–15784. (e) I. Kaneko and T. Hoshino, *J. Org. Chem.* **2016**, *81*, 6657–6671. (f) Y. Terasawa, Y. Sasaki, Y. Yamaguchi, K. Takahashi and T. Hoshino, *Eur. J. Org. Chem.*, **2017**, 287–295.
- [8] (a) S.P.T. Matsuda, L. B. Darr, E. A. Hart, J.B.R. Herrera, K. E. MacCann, M. M. Meyer, J. Pang and H. G. Schepmann, *Org. Lett.*, **2000**, *2*, 2261–2263. (b) B. M. Joubert, L. Hua and S. P. T. Matsuda, *Org. Lett.*, **2000**, *2*, 339–341.

- [9] R. Ito, Y. Masukawa and T. Hoshino, *FEBS J.*, **2013**, *280*, 1267-1280.
- [10] Z-h Lin, H-x Long, Z. Bo, Y-q Wang, and Y-z Wu, *Peptides*, **2008**, *29*, 1798-1805.
- [11] (a) M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel and R. Hilgenfeld, *J. Mol. Biol.* **2001** *307*, 357-377. (b) M. Nishio, Y. Umezawa, J. Fantini, M. S. Weiss and P. Chakrabarti, *Phys. Chem. chem. Phys.*, **2014**, *16*, 124648-12683. (c) D. Pal and P. Chakrabarti, *J. Mol. Biol.*, **1999**, *294*, 271-288.
- [12] M. D. Kolesnikova, W. K. Wilson, D. A. Lynch, A. C. Obermeyer and S. P. T. Matsuda, *Org. Lett.*, **2007**, *9*, 5223-5226.
- [13] (a) T. K. Wu, Y. T. Liu and C. H. Chang, *ChemBioChem*, **2005**, *6*, 1177-1181. (b) T. K. Wu, Y. T. Liu, C. H. Chang, M. T. Yu and H. J. Wang, *J. Am. Chem. Soc.*, **2006**, *128*, 6414-6419.
- [14] 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.
- [15] S. Lodeiro, W. K. Wilson and S. P. T. Matsuda, *Org. Lett.*, **2006**, *8*, 439-442.
- [16] T. Kushiro, M. Shibuya and Y. Ebiduka, *Eur. J. Biochem*, **1998**, *256*, 238-244.
- [17] (a) A. Vigelsø, R. Dybboe, C. N. Hansen, F. Dela, J. W. Helge and A. G. Grau, *J. Appl. Physiol*, **2015**, *118*, 386-394. (b) M. Salmon, R. B. Thimmappa, R. E. Minto, R. E. Melton, R. K. Highes, P. E. O'Maille, A. M. Hemmings and A. Osbourn, *Proc. Natl. Acad. Sci. USA*, **2016**, *113*, E4407-4414.

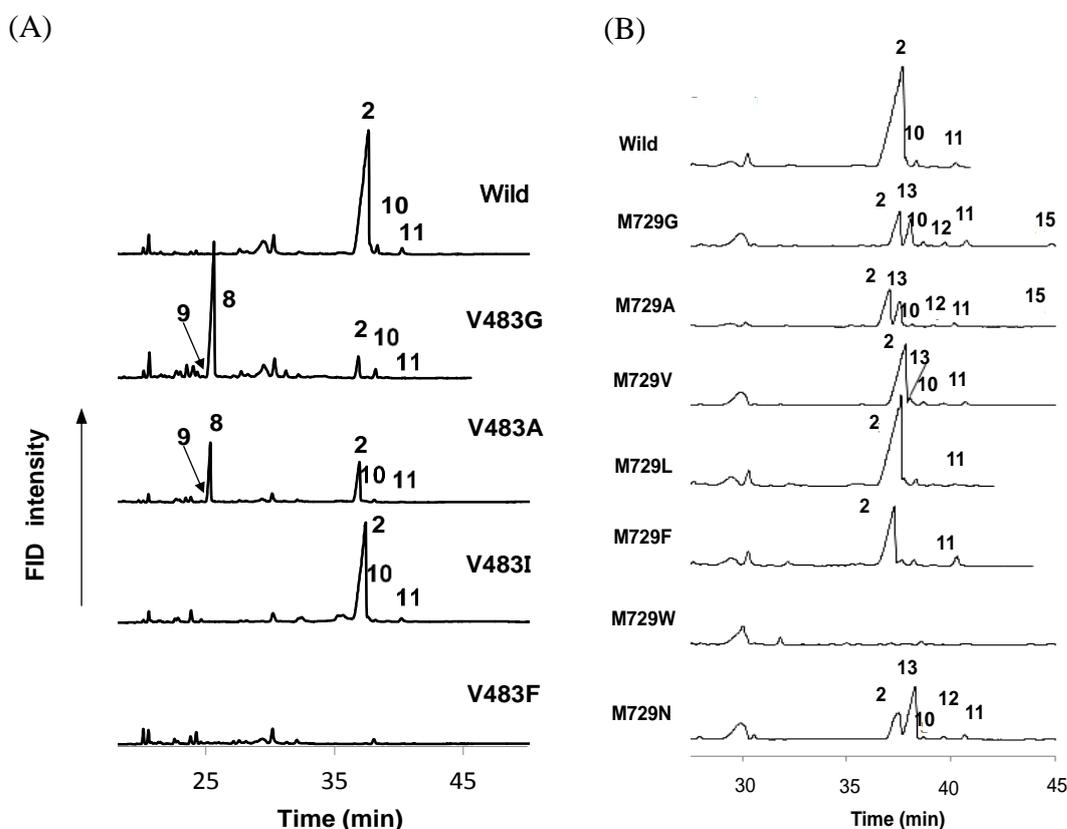
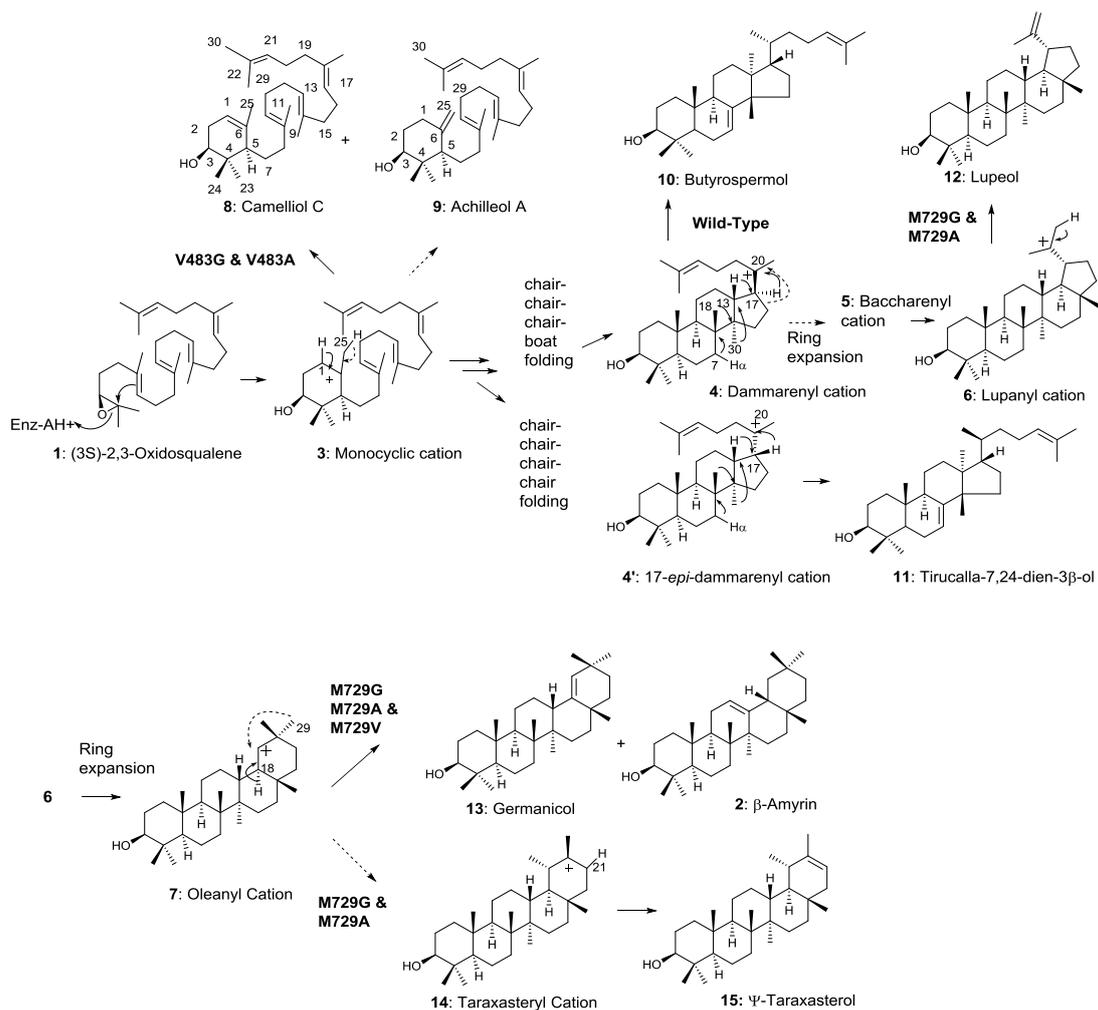


Figure 1. Gas chromatography (GC) profiles of the wild-type *Euphorbia tirucalli* β -amyrin synthase and its V483X (A) and M729X (B) variants. The Gly and Ala variants of Val483 afforded a large amount of monocyclic camelliol C (**8**) and a very small amount of monocyclic achilleol A (**9**). By contrast, the Gly and Ala mutants of Met729 produced germanicol (**13**) in large quantities. The GC conditions for (A) and (B) were as follows: column, J&W, DB-1, capillary (length 30 m, I.D. 0.32 mm, film thickness 0.25 μ m); injection temp., 300°C; column temp., 190–250°C (10°C/min) and 250–268°C (0.35°C/min); injection volume, 3.0 μ L. Products **10–13** and **15** were identified as follows: **10**, butyrospermol; **11**, tirucalla-7,24-dien-3 β -ol; **12**, lupeol; **13**, germanicol; **15**, ψ -taraxasterol.



Scheme 2. Product structures generated by the V483X, W534X, and M729X variants, and the cyclization pathways to forming products 8–13 and 15.

Table 1. Product distribution ratio (%) for V483X mutants.

	Oleanyl cation 7	Monocyclic cation 3		Dammarenyl cation 4	
	β -Amyrin 2	Camelliol C 8	Achilleol A 9	Butyrospermol 10	Tirucalla-7,24-dien-3 β -ol 11
Wild	94.7	—	—	3.1	2.2
V483G	8.3	86.9	4.8	3.5	0.5
V483A	38.1	56.8	1.1	3.4	0.5
V483I	94.0	—	—	3.6	2.4
V483F	—	—	—	—	—

—: not detected.

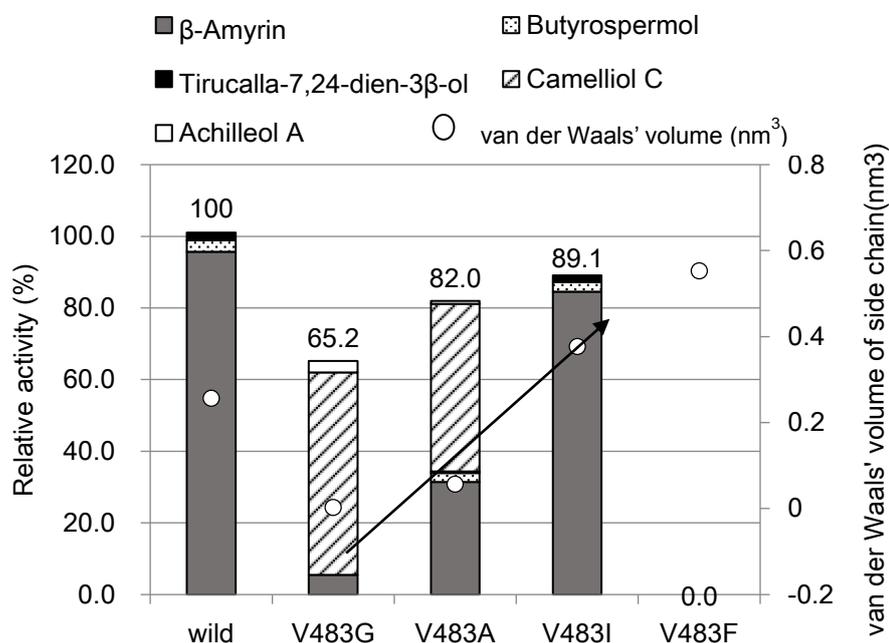


Figure 2 Enzyme activities of V483X variants relative to that of wild-type *Euphorbia tirucalli* β -amyrin synthase. Open circles denote the van der Waals volume (nm^3) of the side chains of the amino acids.^[10] The values were reported to be as follows: Val (wild type), 0.25674; Gly, 0.00279; Ala, 0.05702; Ile, 0.37671; and Phe, 0.55298. The

production of β -amyrin roughly increased in proportion to the increase in the van der Waals volume (see the right-pointing arrow). The relative enzymatic activities of the variants were determined according to the data in Fig. S6.

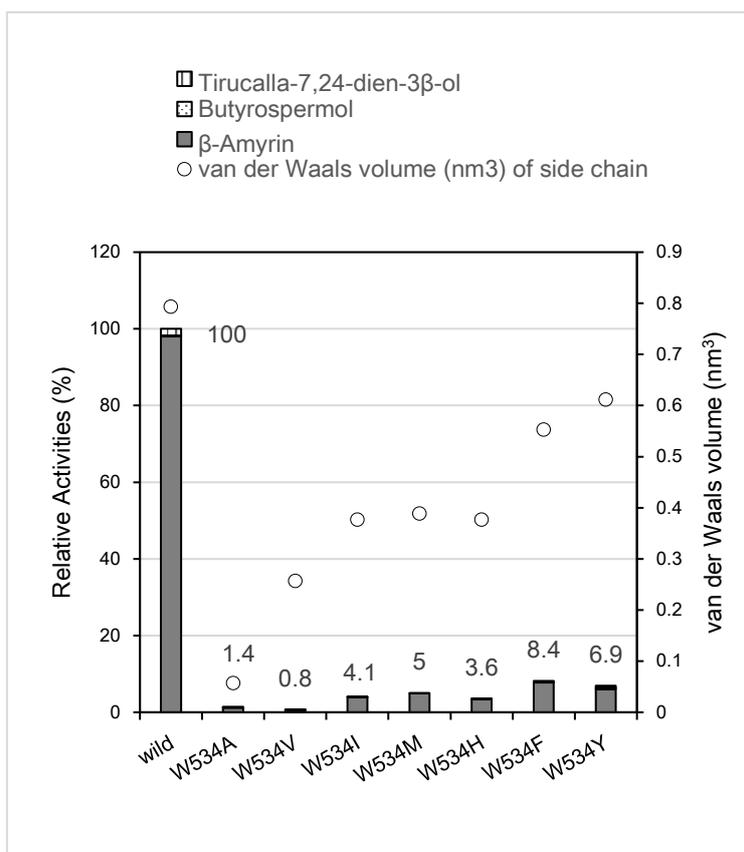


Figure 3. Enzymatic activities of Trp534 variants relative to that of the wild-type *Euphorbia tirucalli* β -amyrin synthase. Significant differences between the aromatic variants (Phe and Tyr mutants) and the aliphatic variants (e.g., Leu and Met mutants) were not observed (see also Fig. S7.2), indicating that the π -electrons at this position are unlikely to be responsible for the polycyclization cascade. The van der Waals volumes were plotted according to literature data.^[10] The relative enzymatic activities of the variants were determined according to the data in Fig. S7.

Table 2. Product distribution ratio (%) for the M729X mutants.

	Oleanyl cation 7			Lupanyl cation 6	Dammarenyl cation 4	
	2	13	15	12	10	11
Wild	95.9	—	—	—	2.0	2.1
M729G	53.5	30.6	2.7	4.1	3.1	6.0
M729A	62.8	27.0	1.1	2.6	1.8	4.7
M729V	88.6	4.8	—	—	3.3	3.3
M729L	98.9	—	—	—	—	1.1
M729F	87.8	—	—	—	3.2	9.0
M729W	—	—	—	—	—	—
M729N	34.5	54.4	-	1.8	6.5	2.8

—: not detected.

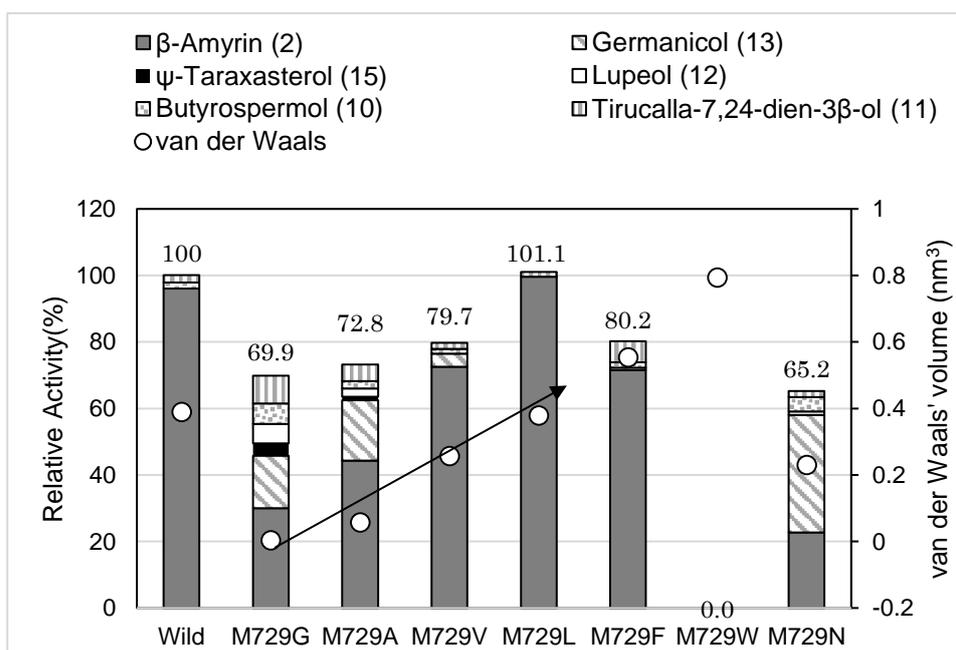


Figure 4. Enzymatic activities of the M729X variants relative to that of wild-type *Euphorbia tirucalli* β -amyrin synthase. Open circles denote the van der Waals volume (nm^3) of the side chains of the amino acids.^[10] The production of β -amyrin was roughly increased in proportion to the increase in the van der Waals volume (see the right-pointing arrow). The relative enzymatic activities of the variants were determined according to the data in Fig. S8.

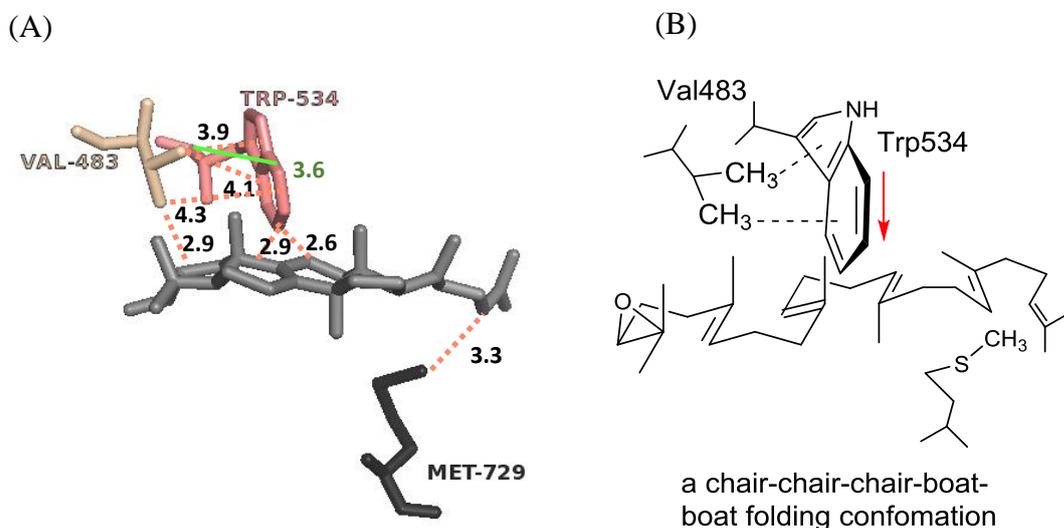


Figure 5. (A) Homology modeling of β -amyrin synthase displayed with PyMOL (<http://www.pymol.org>). This model was constructed by ESyPred3D (<http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/>), based on the X-ray crystal structure of human lanosterol cyclase (PDB ID: 1w6k).^[13] Gray, lanosterol molecule; Black, Met729 residue; Beige, Val483 residue; Pink, Trp534. The numbers indicate distances in Å units. The distance between the carbon atom and the center of the π -system (d_{C-X}) in the EtAS structure are estimated to be 3.9–4.3 Å. Brandl et al. reported that CH- π interaction frequently occurs when the d_{C-X} is ca 3.3–4.4 Å.^[11a] The shortest distance between the methyl group of Val483 and the indole ring (green solid line) are estimated to be 3.6 Å, which is further indicative of CH- π complexation between the side chains, because the shortest distance (3.5–3.9 Å) is also reported as the index of CH- π interaction in protein structures.^[11b, c] The values of 2.6–3.0 Å indicate that the three amino acids are in close proximity to lanosterol or the substrate ligand.

(B) The arrangements of the three active site residues on the folded 2,3-oxidosqualene substrate **1** are illustrated.

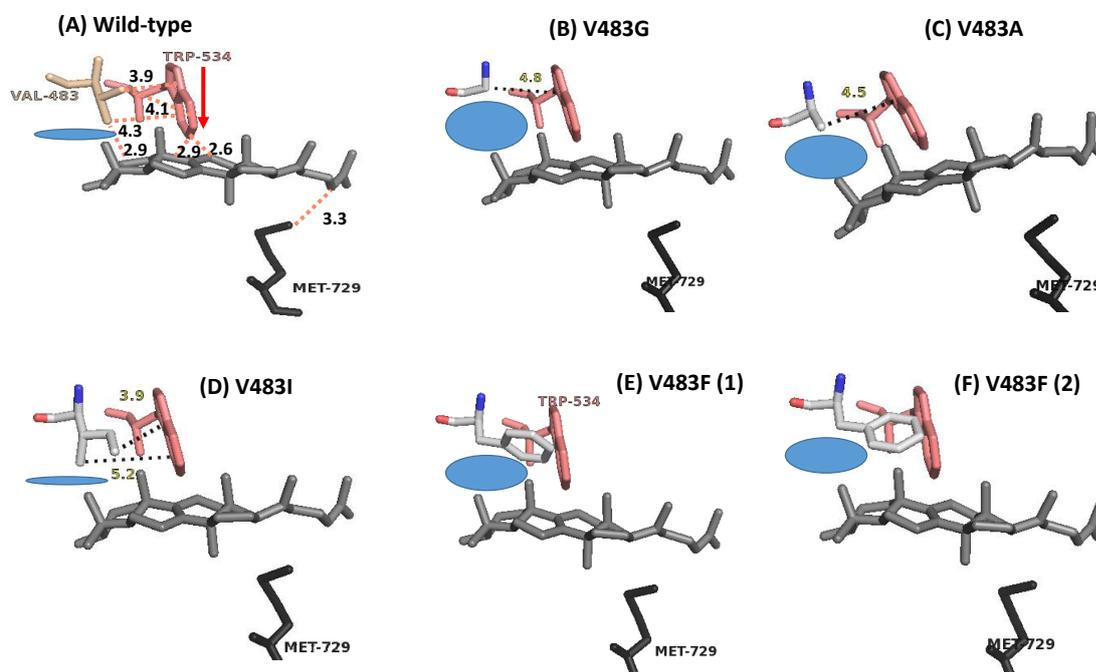


Figure 6. Homology models of the mutated *Euphorbia tirucalli* β -amyrin synthases targeted for the Val483 position. The preferred conformations (rotamers of the side chain) were optimized by PyMOL software. The blue oval shapes display the free space between the mutated residues and the ligand.

- (A) Wild type. CH- π complexation with Trp534. The space between the ligand and Val483 is compact owing to the appropriate steric bulk of the Val residue.
- (B) The V483G mutant. A large space exists between Gly and Trp534, leading to decreased CH- π interaction.
- (C) The V483A variant. A somewhat larger space and decreased CH- π interaction are evident compared with those of the wild-type.
- (D) The V483I variant. Strong CH- π interaction and compactness between the ligand are evident.
- (E) and (F) for the V483F mutant. The orientation of the Phe side chain is different between (E) and (F), which were constructed by minimizing the steric hindrance between the ambient amino acids and Phe. A large space is still generated, and the organized local structure around the A/B-ring formation sites collapses as a result of collision between the side chains of Phe and Trp534, which would probably lead to incorrect positioning of the Trp534 side chains.

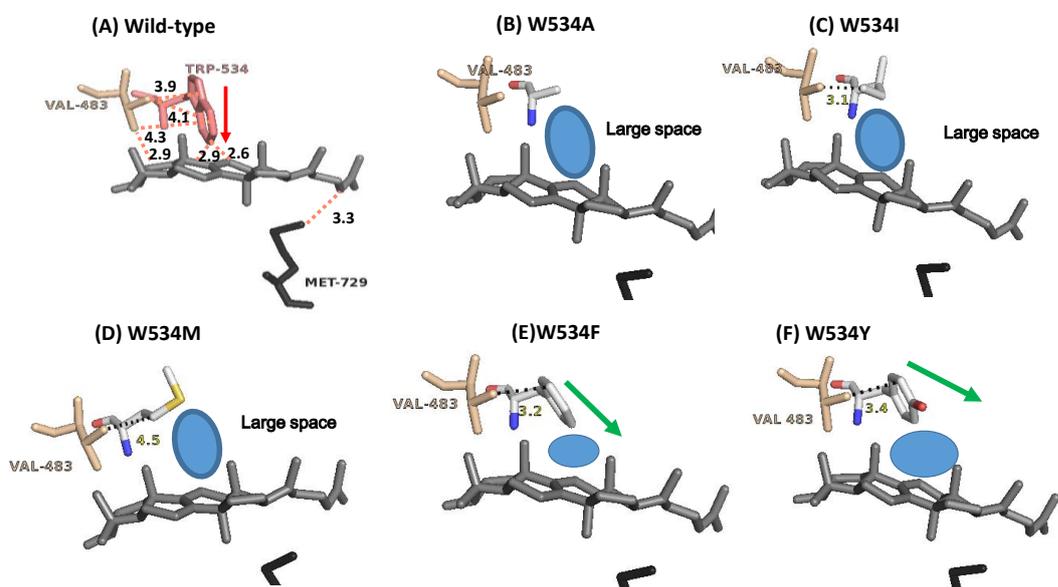
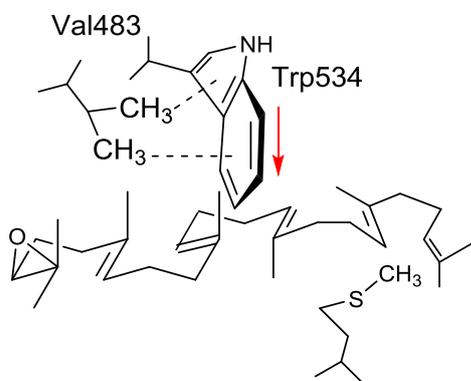


Figure 7. Homology models of the mutated *Euphorbia tirucalli* β -amyrin synthases targeted for the Trp534 position.

The Trp residue is oriented perpendicularly to the ligand around the B/C-ring formation sites (marked with a red arrow). The side chains of the mutated residues are arranged so as to minimize the steric hindrance between the surrounding residues in the cavity. Blue-filled ovals represent the free space between the mutated residues and the ligand. For the aliphatic variants substituted with Ala (B), Ile (C), or Met (D), large vacant spaces are generated between the ligand and the substituents and therefore CH- π interaction does not occur. The green arrows in (E) and (F) display the orientation of the side chains of the Phe and Tyr residues, respectively. These aromatic residues are not arranged perpendicularly to the ligand, which is different from the red arrow (wild-type (A)), although CH- π interactions between the benzene rings and the methyl group of the Val483 residue likely occur because of their close proximity to each other (3.2 or 3.4 Å).

TOC (Table-of-Contents)



a chair-chair-chair-boat-
boat folding conformation

The appropriate steric sizes of the three residues facilitate the folding the substrate into a chair-chair-chair-boat-boat conformation, notably CH- π interaction between the Val and the Trp confers the robust protein architecture around A/B/C-ring formation site.