Article

Identification of Key Amino Acid Residues Determining Product Specificity of 2,3-Oxidosqualene Cyclase in *Siraitia grosvenorii*

Jing Qiao, Jiushi Liu, Jingjing Liao, Zuliang Luo, Xiaojun Ma * and Guoxu Ma *

Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China; qiaojing_happy@126.com (J.Q.); liujiushi90@163.com (J.L.); ljj564293456@126.com (J.L.); zuliangluo@163.com (Z.L.)

* Correspondence: mayixuan10@163.com (X.M.); mgxfl8785@163.com (G.M.);
Tel.: +86-10-5783-3155 (X.M. & G.M.)

Received: 25 October 2018; Accepted: 19 November 2018; Published: 22 November 2018

Abstract: Sterols and triterpenes are structurally diverse bioactive molecules generated through cyclization of linear 2,3-oxidosqualene. Based on carbocationic intermediates generated during the initial substrate preorganization step, oxidosqualene cyclases (OSCs) are roughly segregated into a dammarenyl cation group that predominantly catalyzes triterpenoid precursor products and a protosteryl cation group which mostly generates sterol precursor products. The mechanism of conversion between two scaffolds is not well understood. Previously, we have characterized a promiscuous OSC from *Siraitia grosvenorii* (SgCS) that synthesizes a novel cucurbitane-type triterpene cucurbitadienol as its main product. By integration of homology modeling, molecular docking and site-directed mutagenesis, we discover that five key amino acid residues (Asp486, Cys487, Cys565, Tyr535, and His260) may be responsible for interconversions between chair–boat–chair and chair–chair–chair conformations. The discovery of euphol, dihydrolanosterol, dihydroxyeuphol and tirucallenol unlocks a new path to triterpene diversity in nature. Our findings also reveal mechanistic insights into the cyclization of oxidosqualene into cucurbitane-type and lanostane-type skeletons, and provide a new strategy to identify key residues determining OSC specificity.

Keywords: chemical diversity; oxidosqualene cyclase; homology modeling; site-directed mutagenesis; cucurbitadienol

1. Introduction

Sterols and triterpenes have been shown to exert a wide spectrum of important biological activities [1,2]. They are structurally diverse molecules, over 20,000 steroids and triterpenoids with 200 different skeletons have been found in eukaryotes [3]. These compounds display promising pharmaceutical activity and remarkable stability, as a result, they have attracted much attention from researchers [4,5]. However, naturally producing these ingredients is limited by the difficulties in plant cultivation and artificial extraction; therefore, further exploration of their synthesis and pharmaceutical applications were encouraged. Nevertheless, the challenges of the chemical synthesis of sterols and triterpenes in high yields are still enormous, mainly because of their dramatic structural micro-heterogeneity [6,7].

The important natural products catalyzed by metabolic enzymes from the biosynthetic pathway appears to be a promising alternative, which has attracted growing interest and made progress in the production of triterpenoids and sterols. Metabolic enzymes in the biosynthetic pathway appear to be a promising alternative candidate to catalyze the formation of important natural products,

for this reason, they attracted growing interest and made tremendous strides on the productions of the triterpenoids and sterols. In the process of biosynthesis of all triterpenoids and sterols, the first step is to cyclize a 30-carbon precursor, 2,3-oxidosqualene, that arises from the isoprenoid pathway and is catalyzed by oxidosqualene cyclases (OSCs) [8]. It triggers a complex chain reaction to produce tricyclic, tetracyclic, or pentacyclic molecules. At this point, the divergence of the sterol and triterpenoid biosynthetic pathways depends on different types of OSC. Cyclization of 2,3-oxidosqualene in the chair–boat–chair (CBC) conformation leads to a protosteryl cation intermediate, the precursor of the sterols via the formation of cycloartenol, lanosterol, and parkeol [9,10]. On the contrary, 2,3-oxidosqualene in the chair–chair–chair (CCC) conformation is cyclized into a dammarenyl carbocation intermediate, which subsequently gives rise to diverse triterpenoid skeletons after further rearrangements, see Figure A1. Since then, intensive studies have revealed the function of the OSC enzyme family but its functional diversity is still unclear and further research is needed.

Studies on the cyclization mechanisms go back to the 1950s [11]. Much evidence has shown that the functional diversity of OSCs have been regulated by several key amino acid residues. Up to now, four sterol synthases and two triterpene synthases for the identification and function of the active sites in OSCs have been performed. They are lanosterol synthase from *Homo sapiens* (HsLAS) and *Saccharomyces cerevisiae* (SceERG7) [12–15], squalene-hopene cyclase from *Alicyclobacillus acidocaldarius* (AacSHC) [16], cycloartenol synthase from *Arabidopsis thaliana* (AthCAS) [17–20], β -amyrin synthase from *Avena strigose* (AsbAS) [21], and *Euphorbia tirucalli* (EtAS) [22]. Mutagenesis of key amino acid residues may lead to loss of functions and a change of product types or profiles. Although many mutagenesis studies have focused on the species of *S. cerevisiae* and *A. thaliana*, progress in the study of OSC in medicinal plants has lagged in comparison. Moreover, there has been very little research reported on the key residues which are essential in functional conversion between sterol and triterpenes synthases. Therefore, it is of great significance to search the key amino acids of OSCs, so as to expand their range of products to create more novel structurally diverse products.

Siraitia grosvenorii, a member of Cucurbitaceae family, is a traditional herb mainly planted in the Guangxi Zhuang Autonomous Region in China. Its genome includes two members of the OSC gene family, cucurbitadienol synthase (CS) cyclizes oxidosqualene to form the cucurbitadienol triterpenoid skeleton, which is a distinct step in phytosterol biosynthesis [23,24]. So far, CS has been functionally characterized from *Cucurbita pepo* [25], *Citrullus colocynthis* [26], and *Cucumis sativus* [27]. A previous study suggests that four indirectly interacting residues of CS in *C. sativus* may act collaboratively in the choice of substrate folding type and five-member-ring formation reaction for plant OSCs. Here, using a combined approach that relied on homology modeling, residue coevolution analysis, as well as site-directed mutagenesis, the key amino acid residues and reaction mechanism underlying triterpene product specificity were identified, providing catalytic insights into the functional conversion between sterols and triterpenes.

2. Results

2.1. Catalytic Mechanism Prediction of Cucurbitadienol Biosynthesis in *S. grosvenorii*

Figure A2 shows the homology model of SgCS using the human LaS crystal structure [28] as a template, where the presumed active sites responsible for the construction of the cucurbitadienol scaffold are highlighted. During optimization of the model, no major structural changes were observed for active site residues, and refinement consisted only of fine-tuning the coordinates of the active site sidechain atoms. The active sites of all residues are the same for the SgCS model and LAS, Asp486, Cys487, and Cys565 in SgCS coincide with their counterparts in LAS Asp455, Cys456, and Cys533, respectively.

Molecular docking and energy calculation of quantum chemistry are useful for explaining the experimental results in theory. In order to verify the hypothesis that cucurbitadienol is catalyzed via the C4 intermediate. As shown in Figure A3, the geometry optimization and the single-point

energy calculations were performed on 2,3-oxidosqualene and C-4 cation substrates, respectively. The AutoDock (V4.2) was conducted to dock both 2,3-oxidosqualene and C-4 cation into the active site pocket of SgCS, and the data in Figure 1 indicated that both 2,3-oxidosqualene and C-4 cation could combine with the active site excellently. The computed binding free energy of the C-4 cation (-12.46 kcal/mol) is significantly larger in magnitude than the binding free energy of 2,3-oxidosqualene (-10.86 kcal/mol). A proper position and orientation of the conserved side-chains of His260, Trp413, Phe475, Trp613, and Phe729 was required with the purpose of stabilization of intermediate tertiary cations at C-4. Active site residues Tyr535 possibly participates in CH- π complexation with His260 because of the closer distance (3.3 Å) of the C-4 cation compared to the 2,3-oxidosqualene (4.1 Å). These results clearly indicated that the Tyr535 and His260 together could trigger the 4β -demethylation via oxidative decarboxylation.

In this study, the homology modeling and molecular docking indicated that Asp486, Cys487, Cys565, Tyr535, and His260 in SgCS may be regarded as an active-site residue for the 2,3-oxidosqualene cyclization. These residues were subjected to site-directed mutagenesis experiments to identify the functions of the residues.

Figure 1. The two primary energetically favored binding clusters of the 2,3-oxidosqualene and C-4 cation in the active sites of SgCS from docking studies. (a) 2D docked view of ligand SgCS with 2,3-oxidosqualene receptor; (b) 2D docked view of ligand SgCS with C-4 cation receptor; (c) 2,3-oxidosqualene and protein interaction hydrogen bond interaction view; (d) C-4 cation and protein interaction hydrogen bond interaction view. 3D representation of SgCS is shown in cartoon (gray), the sticks with cyan carbon atoms represent amino acid residues and the sticks with yellow carbon atoms represent 2,3-oxidosqualene and C-4 cation compounds. The red dotted lines denote hydrogen bonds.

2.2. Bioinformatic Analysis of SgCS

The Table A1 has listed the deduced protein parameters that are predicted by the online Pxpasy's ProtParam tool, the SgCS was divided into a stable protein and no signal peptides. Our results suggested that SgCS was mainly found in the nucleus or cytoplasm. The putative protein sequences of SgCS had negative Grand average of hydropathicity (GRAVY) values which were the indication of hydrophilic proteins. Moreover, two transmembrane domains were found in SgCS, including 117–194 AAs and 601–641 AA.

The CS cloned from different species was analyzed using multiple sequence comparison analyses, and the results indicated the presence of the DCTAE motif, which included the catalytic aspartic acid residue at the 486 site and was proposed to be the initiation of the cyclization reaction, see Figure 2 [8]. SgCS also contains three repeats of QW motifs (conserved motifs rich in aromatic amino acids, starting with Q-Gln and ending with W-Trp) which have structure-stabilizing effects and may be useful in the stabilization of the carbocation intermediates during cyclization [29]. In addition, SgCS has a 260-site

histidine residue in the MWCHCR motif, that has been proposed to play an important role in the process of stabilization of the protosteryl cation intermediate [30].

CcCS	AVRLKVGAEVSGVEKEEKVKSISNHLGRQVWFCADQP.....TASPNHLC.....IDNARKHFRNRFHRKQSSDLFLAIQNEKEIA.N	80
CpCS	AVRLKVGAEVSGVEDEEKVKSISNHLGRQVWFCADAAAD.....TPHQLLC.....IQNARNEFHHRFHRKQSSDLFLAIQYKEIA.K	80
CsCS	AVRLKVGAEVSGVEKEEKVKSISNHLGRQVWFCADNDDEDAVIHVVANSKELLQQRQSSSFENARKQFRNRFHRKQSSDLFLTIQYKEIARN	100
SgCS	AVRLKVGAEVSGVEDEEKVKSISNHLGRQVWFCADPAG.....TQQQLLC.....VHKARKAFHDDRFHRKQSSDLFLTIQYKEIVE.N	78
CcCS	CTRGGG.IVKEEEDVRKEITVKNVLERALSFYSAITQNDGNWASDLGGPNLLPGLVIALVITGVLSVLSKHHRQEVACRYLYNHQNEGGWGLHI	179
CpCS	EARGGA.VNKEGEVYKEAVKSTLERALSFYSANQTRDGNWASDLGGPNLLPGLVIALVITGVLSVLSKHHRQEVACRYLYNHQNEGGWGLHI	179
CsCS	CAKNGNTVKEGEVYKEAVKSTLERALSFYSAITQNDGNWASDLGGPNLLPGLVIALVITGVLSVLSKHHRQEVACRYLYNHQNEGGWGLHI	200
SgCS	CGRTAG.VLKEGEVYKEAVKSTLERALSFYSISITQNDGNWASDLGGPNLLPGLVIALVITGVLSVLSKHHRQEVACRYLYNHQNEGGWGLHI	177
CcCS	TVFSGALNYVALRLLEDGCEGGANTKARAWILDRGGATAITSVGKLWLSVLGVYVWSGNNLPPEFWLPPYGLPFHPGRWCHCRVYLPNSVLYGK	279
CpCS	TVFSGALNYVALRLLEDGCEGGANTKARAWILDRGGATAITSVGKLWLSVLGVYVWSGNNLPPEFWLPPYGLPFHPGRWCHCRVYLPNSVLYGK	279
CsCS	TVFSGALNYVALRLLEDGCEGGANTKARAWILDRGGATAITSVGKLWLSVLGVYVWSGNNLPPEFWLPPYGLPFHPGRWCHCRVYLPNSVLYGK	300
SgCS	TVFSGALNYVALRLLEDGCEGGANTKARAWILDRGGATAITSVGKLWLSVLGVYVWSGNNLPPEFWLPPYGLPFHPGRWCHCRVYLPNSVLYGK	274
CcCS	RFVGPITPIVLSLEKELYTIPIYHEIDWNNSRNTCAEDLYYPHPKQDILWGSIVYVEPLFTTRMPGRRLREKALQANSHIHYEDENSRYILGLPVNKKV	379
CpCS	RFVGPITPIVLSLEKELYTIPIYHEIDWNNSRNTCAEDLYYPHPKQDILWGSIVYVEPLFTTRMPGRRLREKALQANSHIHYEDENSRYILGLPVNKKV	379
CsCS	RFVGPITPIVLSLEKELYTIPIYHEIDWNNSRNTCAEDLYYPHPKQDILWGSIVYVEPLFTTRMPGRRLREKAKAKIANSHIHYEDENSRYILGLPVNKKV	400
SgCS	RFVGPITPIVLSLEKELYAVPIYHEIDWNNSRNTCAEDLYYPHPKQDILWGSIVYVEPLFTTRMPGRRLREKALQANSHIHYEDENSRYILGLPVNKKV	374
CcCS	LNNLCCWVEDPYSDAFKFHLQRVVDYLVAEDGNRAQGYNGSLQWDTAFSLQAIISTKLITSFGTILKKAHDFVRKDSQICQEDPDPNNVFRHIIHKGAWP	479
CpCS	LNNLCCWVEDPYSDAFKFHLQRVVDYLVAEDGNRAQGYNGSLQWDTAFSLQAIIVATKLIVSYAPILKKAHDFVRKDSQICQEDPDPNNVFRHIIHKGAWP	479
CsCS	LNNLCCWVEDPYSDAFKFHLQRVVDYLVAEDGNRAQGYNGSLQWDTAFSLQAILSTKLITFTGSLKKAHDFVRKDSQICQEDPDPNNVFRHIIHKGAWP	500
SgCS	LNNLCCWVEDPYSDAFKFHLQRVVDYLVAEDGNRAQGYNGSLQWDTAFSLQAIIVSTKLITFTGSLKKAHDFVRKDSQICQEDPDPNNVFRHIIHKGAWP	474
CcCS	FSTRDHGWLISDCTAEGLKASLNLKLPSTVYGEPLKKNRLCDAVNVLLSLQNDGGFASVELTRSYPWLELINPAETFGDVIIDYDYPVECTSATNEALIT	579
CpCS	FSTRDHGWLISDCTAEGLKASLNLKLPSTVYGEPLKKNRLCDAVNVLLSLQNDGGFASVELTRSYPWLELINPAETFGDVIIDYDYPVECTSATNEALIT	579
CsCS	FSTRDHGWLISDCTAEGLKASLNLKLPSTVYGEPLKKNRLCDAVNVLLSLQNDGGFASVELTRSYPWLELINPAETFGDVIIDYDYPVECTSATNEALA	600
SgCS	FSTRDHGWLISDCTAEGLKASLNLKLPSTVYGEPLERNRLCDAVNVLLSLQNDGGFASVELTRSYPWLELINPAETFGDVIIDYDYPVECTSATNEALIT	574
CcCS	LFKKLHPGERTKEIDTAVAKAANFLENQRTDGSWYGCWGVCFITYAGWFGKGLVAAGRTYSTCAVIRKAGDFLLSKELPGGGWGESVLSQCNKQVYTNLF	679
CpCS	LFKKLHPGERTKEIDTAVAKAANFLENQRTDGSWYGCWGVCFITYAGWFGKGLVAAGRTYSTCAVIRKAGDFLLSKELPGGGWGESVLSQCNKQVYTNLF	679
CsCS	LFKKLHPGERTKEIDTAVAKAANFLENQRTDGSWYGCWGVCFITYAGWFGKGLVAAGRTYSTCAVIRKAGDFLLSKELPGGGWGESVLSQCNKQVYTNLF	700
SgCS	LFKKLHPGERTKEIDTAVAKAANFLENQRTDGSWYGCWGVCFITYAGWFGKGLVAAGRTYSTCAVIRKAGDFLLSKELPGGGWGESVLSQCNKQVYTNLF	674
CcCS	GNNPHLVNTAWLVALLIEAGGAERDPAFLHRAARLLINSQLENGDFPQCEIIVGVFNKNCMIITYAAYRNIFFI WALGEYCHRVLT	763
CpCS	GNNPHLVNTAWLVALLIEAGGAERDPAFLHRAARLLINSQLENGDFPQCEIIVGVFNKNCMIITYAAYRNIFFI WALGEYCHRVLT	763
CsCS	GNNPHLVNTAWLVALLIEAGGAERDPAFLHRAARLLINSQLENGDFPQCEIIVGVFNKNCMIITYAAYRNIFFI WALGEYCHRVLT	784
SgCS	GNNPHLVNTAWLVALLIEAGGAERDPAFLHRAARLLINSQLENGDFPQCEIIVGVFNKNCMIITYAAYRNIFFI WALGEYCHRVLT	758

Figure 2. Multiple sequence alignment of the putative SgCS protein and three other cucurbitadienol synthases, GenBank accession numbers: CcCS (KM821405), CpCS (AB116238), CsCS (AIT72030), SgCS (HQ128567). The highly conserved DCTAE motif marks in red. A terpene synthase feature of QXXXXXW motif is single underlined. MWCHCR motif is shown in the double underline. Important catalytic residues are marked with stars (Asp486, Cys487, Cys565, Tyr535, and His260) which are necessary for 2,3-oxidosqualene cyclization into cucurbitadienol.

2.3. DCTAE Motif for the Initiation of the Polycyclization Cascade

Corey [30] proposed that the D456 residue of the DCTAE motif of *S. cerevisiae* LAS initiates the ring-forming reaction. Asp486 in SgCS coincides with its counterparts in *S. cerevisiae*, LAS Asp456. At first, we constructed three variants (D486N, D486E, D486A), to verify the function of this amino acid residue of SgCS. As shown in Figure 3a, all of the mutants completely lost their function to produce cucurbitadienol because of a single-point mutation at the Asp486 position. This demonstrates that the acidic residue of the Asp is regarded as a proton donor to initiate the cucurbitadienol ring-forming reaction. Interestingly, the D486A and D486N mutants showed multiple profiles of products with a molecular mass of (m/z) 426, whereas D486E produced a minor amount of compound **1** (unidentified) as the sole product, see Figure 3c. Compound **2** was distinguished from an authentic cucurbitadienol standard. The retention times of compounds **3**, **4**, **5**, and **7** were 19.86, 20.71, 22.52, and 25.52 min relative to cucurbitadienol, and were identified as euphol, dihydrolanosterol, dihydroxyeuphol, and tirucallanol, respectively, the data is shown in Figures A4–A7. These additional four compounds can be assigned to different C–C bond-forming conformations. Compound **4** was formed through a chair–boat–chair–chair envelope (C–B–C–C) conformation featuring a trans (CH₃-14 and H-13) orientation at the C–D ring fused carbons to give a 6/6/6/5-fused tetracyclic protosteryl cation. The subsequent reactions of hydride shifts from H-9 to C-8, H-13 to C-17, and H-17 to C-20 together with the methyl shifts from CH₃-8 to C-14 and CH₃-14 to C-13 give the final skeleton of compound **4**. The other compounds, **3** and **7**, initiate the chair–boat–chair (C–C–C) conformation of A/B/C rings as well as the chair and boat of D ring, respectively, to give the damarrenyl cation path. Further hydride and methyl shifts formed compounds **3** and **7**, while compound **5** experienced an additional hydrogenation reaction, see Figure 4.

Based upon the human LAS structure analysis by X-ray crystallography, Thoma [28] proposed that the acidity of D455 of the DCTAE sequence was increased by the hydrogen bond formations with Cys456 and Cys533, versus their counterparts in SgCS with cysteine at position 487 (C487) and cysteine at position 565 (C565), respectively. In order to explore whether additional residues C487 and C485 are included, we constructed six variants (C487M, C487R, C487A, C565M, C565R and C565A), to verify the functions of these two amino acid residues of SgCS. Results showed that the hydrogen-bond with carboxyl residue Asp486 of cys487 is stronger than that with cys565 because the activity of the C487A variant is about 20% of the wild type, whereas the C565A is about 35%, as shown in Figure 3a. The product profiles of 487 and 565 site mutants are summarized in Figure 3b. The Met and Ala mutants produced compounds 3–7 as part of the product profile in addition to cucurbitadienol. On the contrary, the Arg mutants only produced a small number of compound 1.

Figure 3. Product promiscuity of sgCS mutants for cyclization. (a) Cucurbitadienol content in SgCS mutants; (b) product profile of SgCS mutants; (c) extracted ion chromatogram of SgCS and mutants activity. The control panel (pCEV-G₄-K_m empty vector) shows the yeast BY4742 accumulate lanosterol (RT = 15.51 min) in the existence of lanosterol synthase activity. In the presence of SgCS, cucurbitadienol (compound 2) is accumulated.

Figure 4. Oxidosqualene cyclases convert oxidosqualene to diverse cyclic structures.

2.4. Functions of the Tyr535 and His260 Residue

A homology model suggested that the Tyr535 should synergize with the His260 to biosynthesize cucurbitadienol more precisely. Residues His260 and Tyr535 in SgCS correspond to His234 and Tyr510 in Erg7, respectively. Wu [12,13,31] proposed that Tyr510 points to the monocyclic C-10 cation (lanosterol numbering) and that the His234:Tyr510 catalytic base dyad forms the organized local structure of Erg7. A mutant of Y535L of SgCS did not afford any aberrant cyclization products, and only cucurbitadienol was produced in a very small amount. By contrast, the Y535W and Y535A variants of SgCS yielded **3** to **7** in addition to cucurbitadienol. Furthermore, the Trp residue has the largest steric volume among all the proteinogenic amino acids, and thus, the production of sterol compounds **3** to **7** in Y535W is significantly higher than Y535A. Similar to compound **3** to **7**, the content of lanosterol (RT = 15.51 min) in Y535W was obviously increased compared to that of Y535A, see Figure 5. These results demonstrated that substitution of Tyr535 with other amino acids with smaller or larger steric bulks alters the organized local structure, thus influences the polycyclization pathway, resulting in the formation of the products **3** to **7**.

Figure 5. Extracted ion chromatogram of the 535 site and Y535L/C565R double mutant activity.

Next, position 260-directed mutagenesis was used to design more promiscuous OSCs properly, the product profiles of mutants are shown in Table 1. All of the mutants produced compound 5 and 7 as part of the product profile. Interestingly, substitution with acidic amino acid (Asp, Glu), aromatic amino acids (Phe, Trp, Tyr), and hydroxyl amino acid (Ser, Thr) at this position abolish cucurbitadienol biosynthesis, consistent with a specific role for His260 in C4–C8 bond formation. The H260X mutants substituted with hydroxyl amino acids (H260S and H260T) afforded the same product profile including compound 1, 5, and 7. The H260A, H260R, and H260Y mutants produced a minor amount of compound 4, in addition to other products. However, none of the mutants showed a specific product profile. Our data reveal that an aromatic amino acid residue at position 260 is important for promiscuity.

Table 1. Product profile of *S. cerevisiae* BY4742 expressing the SgCS^{H260X} site-saturated mutants.

Amino Acids Substitution	Product Profile						
	1	2	3	4	5	6	7
Ala (A)		18	19	5	37	5	16
Cys (C)		43			24		32
Asp (D)			20		59		21
Glu (E)			43		34		23
Phe (F)					53		47
Gly (G)		10	18		52		20
Ile (I)		24	36		23		17
Lys (K)					58	9	33
Leu (L)		11	32		31	4	21
Met (M)		43	29		17		11
Asn (N)		48	9		32	2	8
Pro (P)		15	21		22	18	24
Gln (Q)		71	3		17		9
Arg (R)			20	6	49	4	22
Ser (S)	8				56		36
Thr (T)	10				55		35
Val (V)		25			40	13	22
Trp (W)					46	24	30
Tyr (Y)				20	52		28

2.5. Double Mutants Determining Product Specificity

Various possible SgCS mutants were created via replacing amino acid residues at each of these five sites in combination and expressing them in BY4742. The D486E/C487R double mutants obtained an ability to specifically produce compound **1**, while the production was not increased in comparison with single mutant D486E or C487R. The D486A/C487A double mutants also didn't change the product profile compared to the D486A or C487A single mutant. Interestingly, compound **6** synthesis was significantly increased in the double mutant Y535L/C565R, producing at least up to 15 times more product than that of the single mutant, see Figure 5. These results showed that residues may develop synergism in the choice of an extraordinary substrate folding intermediate for OSCs.

2.6. Reaction Mechanism of SgCS

A one-step reaction to form the tetracyclic cucurbitadienol molecule liner squalene is considered to be a very complex biochemical reaction in nature. Based on the results of our investigation, we proposed a hypothetical mechanism of the SgCS-catalyzed 2,3-oxidosqualene cyclization reaction of cucurbitadienol, shown in Figure 6. Firstly, the cyclization reaction is initiated by Asp486 protonating the epoxide group of prefolded 2,3-oxidosqualene. Cys487 and Cys565 serve as hydrogen-bonding partners with Asp486; therefore, there is an increased acidity of Asp486 in SgCS. Then, the OSC cyclization cascade stops with the tertiary protosterol cation at C-20 when the formation of the five-membered D-ring is accomplished. Secondly, the protosterol C-20 cation is transformed to the cucurbitadienol C-4/C-8 cation via a rearranging skeleton. Ultimately, the highly conserved His 260 is the basic residue that is close enough to accept the proton in the specific deprotonation of the C-4/C-8 cucurbitadienol cation that terminates catalysis. The hydroxy group of Tyr535 that is hydrogen-bonded to His 260 would be in an appropriate position for the final deprotonation step.

Figure 6. A proposed tentative mechanism for the SgCS-catalyzed cyclization reaction.

3. Discussion

We identified five key amino acid positions which regulate the functional switch between cucurbitane-type and lanostane-type triterpenoids synthesis using structural modeling. The use of protein homology modeling combined with molecular docking is a useful method to discover new enzymes, and it reveals a chemical diversity of catalytic mechanisms of enzymes [32]. In the present study, this method enabled us to identify the novel phytosterols, euphol, dihydroxyeuphol, and tirucallenol, and uncover an unknown field regarding OSC cyclization. As far as we know, biosynthesis of these compounds has not been reported before, their structures are quite similar to cycloartenol's and lanosterol's, that are the precursors of sterol biosynthesis. The products of SgCS mutants may participate in synthesizing new steroid hormones, and, therefore, affect the growth and development of a plant.

A recent research of bacterial OSCs (HsLAS) has found four sites at positions Trp230, His232, Tyr503, and Asn697 that may lead to a change in product from tetracyclic triterpenoid lansterol to pentacyclic products, but keep a protosteryl-type C-B-C conformation [33]. Two residues, at position Phe696 and Ser699, were identified in the studies of *Avena strigosa* and *Euphorbia tirucalli* b-amyrin synthase (SAD1), which determined the functional interconversion of b-amyrin and lansterol synthases. These enzymes produce pentacyclic and tetracyclic triterpenes from the same dammarane-type C-C-C conformation [21,34]. Most of the tetracyclic products catalyzed via the C-B-C intermediate are precursors for primary metabolites, such as cholesterol and brassinosteroids (BRs) [35]. Although, one exception is cucurbitadienol synthase from cucurbit plants [25,27]. In this study, the five key residue positions in SgCS, Asp486, Cys487, Cys565, Tyr535, and His260, are responsible for the alternation between the C-B-C and C-C-C conformations. This is the first enzyme to report the conformational change from the C-C-C to the C-B-C folding conformation among the OSC mutagenesis studies reported hitherto.

Over the course of evolution, in order to adapt to both natural and artificial selections, the plants produce diverse catalogs of compounds via improvements to their system. Research on engineering enzymes has progressed, such as a broadened substrate, reaction, and product specificity. In this study, because of a single-point mutation, the product scope of SgCS for oxisqualene-cyclization was significantly enlarged. It has proven to be dramatically more difficult to redesign enzymes to have a stringent specificity. There are various factors that can be used to improve the product specificity and stereoselectivity. During the cyclization process, specific cation- π interactions mediated by aromatic residues and cation intermediates are important for OSC's generation of the specific final product [20,36–38]. It is worthy to note that we found two residues, Cys565 and Tyr535, can act collaboratively in the choice of the specific substrate folding intermediate for OSCs. These results offered a possibility that may produce promiscuous OSC variants to create specific products.

Since the 1950s, the response mechanism of OSC has been studied in depth. Many OSCs have been cloned and sequenced since the first DNA sequence of *Candida albicans* OSC was reported in 1990 [39]. However, isolating OSCs in an active form and characterizing their enzymatic properties in vitro has been very difficult because they are membrane-bound proteins. To date, only the X-ray crystallographic structures of the two triterpene cyclases of SHC and human lanosterol cyclase have been solved and the structure of cucurbitadienol synthase has not yet been elucidated. Recently, a report describing the functional analyses of the active sites in cucurbitadienol synthase have appeared, but at present, studies on the catalytic mechanism of cucurbitadienol synthase remain insufficient, and further research is needed to better understand cucurbitadienol biosynthesis [40]. To acquire detailed information on the polycyclization cascades promoted by triterpene cyclases, collaborations involving multiple research fields, such as molecular biology, structural biology, biochemistry, and organic chemistry, will be required.

4. Materials and Methods

4.1. Chemicals and General Methods

The host strain *S. cerevisiae* BY4742 was purchased from Invitrogen (Carlsbad, CA, USA). Chemical reagents were purchased from Sigma Aldrich (St. Louis, MO, USA), J&K Scientific Ltd. (Beijing, China), and Taihe Biotechnology Co. Ltd. (Beijing, China). KOD-Plus DNA Polymerase was purchased from TOYOBO Biotech Co. Ltd. (Shanghai, China). Primers synthesis and DNA sequencing were obtained from Sangon Biotech Company (Beijing, China). Restriction enzymes and DNA ligase were purchased from Takara Biotechnology Co. Ltd. (Dalian, China). The yeast expression plasmids pCEV-G4-Km (Addgene plasmid # 46819) were a gift from Lars Nielsen and Claudia Vickers [41]. Yeast products were detected using an Acquity UPLC™ system (Waters Corp. USA) with an Acquity BEH C18 column (100 mm × 2.1 mm i.d. 1.7 µm), and purified by a Lumiere K-1001 pump, a Lumiere K-2501 single λ absorbance detector and a YMC-Pack ODS-A column (5 µm, 10 × 250 mm, YMC, Kyoto, Japan). ESI-MS was performed on a LTQ-Orbitrap XL spectrometer (Thermo Fisher Scientific, Boston, MA, USA). The conversion rates of the substrate were calculated from peak areas of products as analyzed by UPLC at their maximum absorption wavelength, respectively. Compounds were characterized by ¹H NMR at 600 MHz and ¹³C APT at 150 MHz on Bruker AV III 600 NMR spectrometers. Chemical shifts (δ) were referenced to internal solvent resonances and were given in parts per million (ppm). Coupling constants (J) were given in hertz (Hz).

4.2. Molecular Modeling

The comparison of different amino acid sequences using alignment was carried out by ClustalX (<http://www.clustal.org>). The sequence alignments were edited and analyzed using ESPript 3.0 (<http://escript.ibcp.fr/ESPript/ESPript/>).

The three-dimensional structure models of SgCS were constructed using Modeller 9.17 (Accelrys, Inc., San Diego, CA, USA). The human lanosterol synthase (PDB accession code 1W6J) was used as the template for modeling. The generated three-dimensional structures were improved by using the loop refined program of Modeller 9.17. The final models were checked by Procheck (<http://nihserver.mbi.ucla.edu/511SAVES/>) and 90.4% of residues were located in the most favored regions in the Ramachandran plot.

4.3. Molecular Docking

AutoDock (V4.2) was used with an empirical free-energy function to evaluate binding free energies and the Lamarckian genetic algorithm (LGA) to search for favorable binding positions [42]. The empirical scoring function which contains hydrogen bonding, electrostatics, conform, torsion, and solvent terms, was trained to calculate the affinity between ligand and protein [43]. It was used to dock compounds into the SgCS protein. The AutoGrid was used to calculate the grid maps which can define the search region and represent the protein during the docking process, the dimensions of grid maps were 50 Å × 50 Å × 40 Å centered on coordinates 26.271, 68.084, and 7.604, with a spacing of 0.375 Å between the grid points. The LGA parameters were accepted as number of GA runs 100, population size 150, maximum number of evals 2,500,000 generations and other parameters were left at the default values.

4.4. Mutagenesis Experiments of SgCS

The CDS of SgCS biosynthetic gene (GenBank accession No. HQ128567) was codon-optimized (Box A1) for synthesis and cloned into the *Bam*HI/*Eco*RI sites of the pCEV-G4-Km-USER [44] yeast expression vector under the control of the TEF1 promoter to construct pCEV-G4-Km-USER-SgCS. The Site-Directed Mutagenesis Kit obtained from Biomed (Beijing, China) was used to detect the mutagenesis, and the corresponding degenerate primers are given in Table A2 with the substitutions underlined. After separation by 1% agarose gel electrophoresis, the resulting PCR products were

directly transformed into the *Trans1-T1 E. coli*. Sanger sequencing was used to determine the sequence of the mutation in the resulting plasmid (pCEV-G4-Km) using the oligonucleotide primers pCEV-Seq, the data is shown in Table A2.

4.5. Yeast Transformation and Cell Cultivation

The plasmids were transfected into *S. cerevisiae* strains BY4742 using the Frozen-EZ yeast transformation II kit obtained from Zymo Research (Orange, CA, USA) and selected for growth on YPD plates with 200 mg/L G418. The empty pCEV-G4-Km vector was also introduced into BY4742 as a control.

The recombinant cells were transferred and cultured in 2 mL YPD medium with 200 mg/L G418 and in a humidified incubator at 30 °C, and 250 rpm to 600 nm (OD600) of approximately 1.0. Then flasks (250 mL) containing 100 mL medium were then inoculated to an OD600 of 0.05 with the seed cultures. After inoculation, incubate the strains at 30 °C, 250 rpm for 3 days. Moreover, Shimadzu UV-2550 spectrophotometer was used to detect the optical densities at OD600.

4.6. UPLC-Q-TOF Analysis of Yeast Extracts

Yeast cells were harvest from culture using a centrifuge at $10,000 \times g$ for 5 min, then washed with 5 mL of 20% KOH/50% ethanol and extracted three times with the same volume of n-hexane. The recombinant extracts were distilled and dried under reduced pressure and dissolved in 1 mL of acetonitrile.

An Acquity UPLCTM system was used to detect chromatography using a BEH rapid resolution RP-C18 column (1.7 μ m, 2.1 $\times$ 100 mm, Waters). The mobile phase consisted of 0.05% formic acid aqueous solution (*v/v*, solvent A) and acetonitrile (solvent B) with a gradient elution: 0–50 min, 90–100.0% (B); 50–60 min, 100.0% (B). The flow was 0.3 mL/min and the injection volume was 3 μ L.

4.7. Compound Purification from Yeast

Yeast harboring 535W SgCS mutant was cultured (50 L) and then the mixture was centrifuged at 3000 rpm for 10 min, the pellet (600 g) was resuspended in 3 L 20% KOH/50% EtOH (1:1), and treated with an alkali lysis method for 1 h at 95 °C, and then double extraction of with the same volume of n-hexane. After retrieval of the n-hexane, the residue (2.4 g) was subjected to a column chromatograph (CC) on a silica gel (200–300 mesh, 3 $\times$ 10 cm) with a gradient of petroleum ether-CH₂Cl₂ (6:1–1:2) to yield seven fractions (Fr.1–Fr.7). Fr.3 (2:1, 0.26 g) was purified by using semi-preparative HPLC eluting with CH₃CN-H₂O (98:2) on a YMC RP C-18 column to give compounds **3** (*R*_t 65.0 min), **4** (*R*_t 73.3 min), **5** (*R*_t 77.7 min), **7** (*R*_t 102.6 min). The flow rate was kept constant at 6 mL/min and the detection wavelength was set at 203 nm.

4.8. Compound Structure Elucidation by NMR Data Analysis

NMR analyses of compounds **3–5**, and **7** were performed on a BrukerAvance III 600 analysis system with the solvent of CDCl₃, and the structures were elucidated according to optical rotation, ESIMS, and NMR data (¹H NMR, ¹³C APT, and NOESY spectra) and compared with the reported ones.

Product **3**: [α]_D²⁰ = +32.4 (*c* 0.10, CHCl₃); ESIMS: *m/z* 449 [M + Na]⁺; ¹H NMR (600 MHz, CDCl₃) δ _H: 5.10 (1H, t, *J* = 7.2 Hz, H-24), 3.22 (1H, dd, *J* = 12.0, 4.8 Hz, H-3), 0.67 (3H, s, H-18), 0.81 (3H, s, H-29), 0.86 (3H, s, H-30), 0.91 (3H, d, *J* = 7.2 Hz, H-21), 0.97 (3H, s, H-28), 0.99 (3H, s, H-19), 1.28 (3H, s, H-26), 1.66 (3H, s, H-27); ¹³C NMR (150 MHz, CDCl₃) δ _C: 35.6 (C-1), 27.9 (C-2), 79.0 (C-3), 38.9 (C-4), 50.1 (C-5), 18.3 (C-6), 28.2 (C-7), 134.3 (C-8), 130.9 (C-9), 37.0 (C-10), 21.0 (C-11), 26.5 (C-12), 44.4 (C-13), 49.8 (C-14), 31.0 (C-15), 30.8 (C-16), 50.4 (C-17), 15.8 (C-18), 19.2 (C-19), 36.2 (C-20), 18.7 (C-21), 36.4 (C-22), 24.9 (C-23), 125.3 (C-24), 131.0 (C-25), 18.6 (C-26), 25.8 (C-27), 24.3 (C-28), 15.4 (C-29), 28.0 (C-30).

Product **4**: [α]_D²⁰ = +62.2 (*c* 0.10, CHCl₃); ESIMS: *m/z* 447 [M + Na]⁺; ¹H NMR (600 MHz, CDCl₃) δ _H: 3.25 (1H, dd, *J* = 12.0, 1.8 Hz, H-3), 0.81 (3H, s, H-18), 0.98 (3H, s, H-19), 0.88 (3H, d, *J* = 7.2 Hz, H-21),

0.86 (3H, s, H-28), 0.92 (3H, d, $J = 6.0$ Hz, H-26), 0.93 (3H, d, $J = 6.0$ Hz, H-27), 1.01 (3H, s, H-29), 0.70 (3H, s, H-30); ^1H NMR (150 MHz, CDCl_3) δ_{C} : 36.1 (C-1), 27.9 (C-2), 79.0 (C-3), 38.9 (C-4), 50.9 (C-5), 21.0 (C-6), 28.2 (C-7), 134.3 (C-8), 130.9 (C-9), 37.0 (C-10), 18.3 (C-11), 25.5 (C-12), 44.4 (C-13), 49.8 (C-14), 31.0 (C-15), 29.7 (C-16), 50.4 (C-17), 18.8 (C-18), 18.6 (C-19), 36.2 (C-20), 19.1 (C-21), 35.6 (C-22), 31.3 (C-23), 25.6 (C-24), 31.3 (C-25), 24.2 (C-26), 24.8 (C-27), 15.4 (C-28), 28.0 (C-29), 24.3 (C-30).

Product 5: $[\alpha]_{\text{D}}^{20} = +33.7$ (c 0.10, CHCl_3); ESIMS: m/z 447 $[\text{M} + \text{Na}]^+$; ^1H NMR (600 MHz, CDCl_3) δ_{H} : 3.23 (1H, dd, $J = 12.0, 1.8$ Hz, H-3), 0.68 (3H, s, H-18), 0.81 (3H, s, H-29), 0.86 (3H, s, H-30), 0.90 (3H, d, $J = 7.2$ Hz, H-21), 0.97 (3H, s, H-28), 1.00 (3H, s, H-19), 0.87 (3H, d, $J = 6.0$ Hz, H-26), 0.89 (3H, d, $J = 6.0$ Hz, H-27); ^1H NMR (150 MHz, CDCl_3) δ_{C} : 35.5 (C-1), 27.9 (C-2), 79.0 (C-3), 38.9 (C-4), 50.1 (C-5), 18.4 (C-6), 28.2 (C-7), 134.4 (C-8), 130.9 (C-9), 37.0 (C-10), 21.1 (C-11), 26.3 (C-12), 44.4 (C-13), 49.8 (C-14), 31.0 (C-15), 30.5 (C-16), 50.4 (C-17), 15.8 (C-18), 19.1 (C-19), 36.2 (C-20), 18.9 (C-21), 36.2 (C-22), 31.2 (C-23), 25.5 (C-24), 31.1 (C-25), 24.2 (C-26), 25.7 (C-27), 24.3 (C-28), 15.4 (C-29), 27.9 (C-30).

Product 7: $[\alpha]_{\text{D}}^{20} = +11.7$ (c 0.10, CHCl_3); ESIMS: m/z 449 $[\text{M} + \text{Na}]^+$; ^1H NMR (600 MHz, CDCl_3) δ_{H} : 5.08 (1H, t, $J = 7.2$ Hz, H-24), 3.25 (1H, dd, $J = 12.0, 4.8$ Hz, H-3), 0.68 (3H, s, H-18), 0.81 (3H, s, H-29), 0.86 (3H, s, H-30), 0.85 (3H, d, $J = 7.2$ Hz, H-21), 0.97 (3H, s, H-28), 0.99 (3H, s, H-19), 1.28 (3H, s, H-26), 1.66 (3H, s, H-27); ^1H NMR (150 MHz, CDCl_3) δ_{C} : 35.6 (C-1), 27.8 (C-2), 79.0 (C-3), 38.9 (C-4), 50.1 (C-5), 18.4 (C-6), 28.2 (C-7), 134.3 (C-8), 130.9 (C-9), 37.1 (C-10), 21.0 (C-11), 26.5 (C-12), 44.5 (C-13), 49.7 (C-14), 31.0 (C-15), 30.8 (C-16), 50.4 (C-17), 15.8 (C-18), 19.2 (C-19), 36.2 (C-20), 24.1 (C-21), 36.4 (C-22), 24.9 (C-23), 125.1 (C-24), 131.0 (C-25), 18.6 (C-26), 25.8 (C-27), 24.3 (C-28), 15.4 (C-29), 28.0 (C-30).

5. Conclusions

To summarize, we developed a reliable homology model of SgCS to explore the potential function of the active site. Molecular docking and energy calculation of quantum chemistry were useful for explaining the catalytic mechanism in theory. The dynamic domain of SgCS was constituted by five covarying residues including Asp486, Cys487, Cys565, Tyr535, and His260. We described a site-directed mutagenesis strategy for the synthesis of cucurbitane-type and lanostane-type derivatives via an oxidosqualene cyclization reaction using SgCS as a safe, economical, and eco-friendly catalyst. The results suggested that the active site of SgCS accounted for the alternation between C–B–C and C–C–C conformations. A lot of useful information about enzyme evolution can be provided by exploring the untapped catalytic promiscuity of natural enzymes. OSCs can polycyclize compounds in a one-step reaction, as a result, it is currently being applied as an interesting biocatalyst in a variety of biotechnological processes. Its tight stereo control, when compared to chemically produced compounds, makes this enzyme greatly superior to catalyze the reaction. Further studies are needed to broaden the substrate range of the enzyme, which makes OSCs a promising tool to produce novel, unnatural cyclic products.

Author Contributions: Conceptualization, X.M. and J.Q.; Methodology, Z.L., J.L. (Jingjing Liao) and G.M.; Formal Analysis, G.M.; Data Curation, J.Q. and Z.L.; Writing-Original Draft Preparation, J.Q. and G.M.; Writing-Review & Editing, X.M.; Manuscript Review, J.L. (Jiushi Liu); Project Administration, G.M. and X.M.; Funding Acquisition, X.M.

Funding: This work was supported by the National Natural Science Foundation of China (NO. 81573521), Beijing Municipal Natural Science Foundation (NO. 5172028) and CAMS Innovation Fund for Medical Sciences (NO. 2017-I2 M-1-013).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Box A1. Code-Optimized SgCS gene sequence.

GGATCC GCCACC
 ATGTGGAGATTGAAAGTTGGTGCTGAATCTGTTGGTGAAAATGATGAAAAATGGTTGAAATCTA
 TCTCTAACCACCTGGGTAGACAAGTTTGGGAATTTTGTCCAGATGCTGGTACTCAACAACAATTG
 TTGCAAGTTTCATAAAGCTAGAAAAGCTTTTCATGATGATAGATTCCATAGAAAACAGTCTTCTGA
 TTTGTTTATCACTATCCAATACGGTAAAGAAGTTGAAAACGGTGGTAAAACTGCTGGTGTTAAAT
 TGAAAGAAGGTGAAGAAGTTAGAAAAGAAGCTGTTGAATCTTCTTTGGAAAGAGCTTTGTCTTT
 TTATTCTTCTATCCAAACTTCTGATGGTAACTGGGCTTCTGATTGGGTGGTCCAATGTTTTGTG
 CCAGGTTTGGTTATTGCTTTGTATGTTACTGGTGTTTTGAATTCTGTTTTGTCTAAGCATCATAGAC
 AAGAAATGTGTAGATATGTTTATAATCACCAGAACGAAGATGGTGGTTGGGGTTTGCATATTGAA
 GGTCCATCTACTATGTTTGGTTCTGCTTTGAATTATGTTGCTTTGAGATTGTTGGGTGAAGATGCTA
 ATGCTGGTGCTATGCCAAAAGCTAGAGCTTGGATTTTGGATCATGGTGGTGCTACTGGTATTACTT
 CTGGGGTAAATTGTGGTTGTCTGTTTTGGGTGTTTATGAATGGTCTGGTAATAATCCATTGCCAC
 CAGAATTTTGGTTGTTTCCATATTTTTGCCATTTTCATCCAGGTAGAATGTGGTGTCAATTGTAGAAT
 GGTTTATTTGCCAATGTCTTATTTGTACGGTAAAAGATTTGTGGTCCAATTACTCCAATTGTTTTG
 TCTTTGAGAAAAGAATTGTATGCTGTTCCATATCATGAAATCGATTGGAATAAATCTAGAAACAC
 TTGCGCTAAAGAAGATTTGTATTACCCACATCCAAAATGCAAGATATTTGTGGGGTCTTTTGC
 ATCATGTTTATGAACCATTGTTTACTAGATGGCCAGCTAAAAGATTGAGAGAAAAAGCTTTGCAA
 ACTGCTATGCAACATATTCATTATGAAGATGAAAATACTAGATATATTTGTTTGGGTCCAGTTAAC
 AAGGTTTTGAATTTGTTATGTTGTTGGGTGAAGATCCATATTCTGATGCTTTTAAATTACATTTAC
 AAAGAGTTCATGATTATTTATGGGTTGCTGAAGATGGTATGAAAATGCAAGGTTATAATGGTTCT
 CAATTATGGGATACGCTTTTTCTATCCAAGCTATTGTTTCTACTAAGTTAGTTGACAATTACGGTC
 CTACTTTAAGAAAGGCTCATGATTTTGTAAAGTCTTCTCAAATCCAACAAGATTGCTCTGGTGATC
 CTAATGTTTGGTATAGACACATTCACAAGGGTGCTTGGCCTTTTTCTACTAGAGATCACGGTTGGT
 TAATTTCTGACTGACTGCTGAAGGTTTAAAGGCTGCTTTAATGTTATCTAAGTTACCTTCTGAAA
 CTGTTGGTGAATCTTTAGAAAGAAATAGATTATGCGACGCAGTTAATGTTTTATTATCTTTACAGA
 ACGACAATGGTGGTTTCGCATCTTATGAATTAACAAGATCTTACCCTTGGTTAGAATTAATTAATC
 CTGCAGAAACATTCGGTGACATGTTATTGACTACCCCTTACGTTGAGTGATACATCTGCAACAATG
 GAGGCATTAACATTATTCAAGAAGTTACACCCTGGTCACAGAACAAAGGAGATAGACACAGCA
 ATAGTTAGAGCAGCAATTTCTTAGAGAATATGCAAAGAACAGACGGTCTTGGTACGGTTGTTG
 GGGTGTGTTGTTTACATACGCAGGTTGGTTCCGTATAAAGGGTTAGTTGCAGCAGGTAGAACAT
 ACAATAATTGCTTAGCAATAAGAAAGGCATGCGACTTCTTATTATCTAAGGAGTTACCTGGTGGT
 GGTTGGGGTGAGTCTTACTTATCTTGCCAGAATAAGGTTTACACAACTTAGAGGGTAAACAGAC
 CTCACTTAGTTAACACAGCATGGGTTTTAATGGCATTAAATAGAGGCAGGTACGGCAGAGAGAGA
 CCCTACACCTTTACACAGAGCAGCAAGATTATTAATAAACTCTCAGTTAGAGAACGGTGACTTCC
 CTCAGCAGGAGATAATGGGTGTTTTCAACAAGAACTGCATGATAACATACGCAGCATACAGAAA
 CATATTCCTATCTGGGCATTAGGTGAGTACTGCCACAGAGTTTTAACAGAG
GAATTC
 Nucleotides indicating restriction sites are underlined and bold.

Table A1. Parameters computed by the online ExPASy Proteomics Server.

Protein Characteristic	SgCS
Number of the deduced amino acids	759
Molecular weight (kDa)	86.4
Theoretical pI	6.64
Instability index	38.07
Grand average of hydropathicity (GRAVY)	−0.313

Table A2. List of primer sequences.

Mutants	Primers (5' → 3')
USER cassette oligo for pCEV-G4-Km, SpeI/PacI overhang.	F: CTAGTCTGAGGCATGTTTAAACGACCCTCAGCTTAAT R: TAAGCTGAGGGTCGTTTAAACATGCCTCAGCA
SgCS-Wild	F: GGCATGTTUATGTGGAGATTGAAAGTTGGTGC R: GGGTCGTTUTCACTCTGTAAAACTCTGTGGCAGTAC
pCEV-Seq	F: CGATGACCTCCCATGATA R: CGTTCTTAATACTAACATAACT
D486E	F: AGATCACGGTTGGTTAATTTCT <u>GAA</u> CTACTGCTGAAGGTTTAAAGGC R: GCCTTTAAACCTTCAGCAGTAC <u>TTC</u> AGAAATTAACCAACCGTGATCT

Table A2. Cont.

Mutants	Primers (5'→3')
D486N	F: AGATCACGGTTGGTTAATTTCTA <u>ACG</u> TACTGCTGAAGGTTTAAAGGC R: GCCTTTAAACCTTCAGCAGTAC <u>GTT</u> AGAAATTAACCAACCGTGATCT
D486A	F: AGATCACGGTTGGTTAATTTCT <u>GCT</u> TACTGCTGAAGGTTTAAAGGC R: GCCTTTAAACCTTCAGCAGTAC <u>AGC</u> AGAAATTAACCAACCGTGATCT
C487M	F: GACTACCCTTACGTTGAGAT <u>G</u> ACATCTGCAACAATGGAGG R: CCTCCATTGTTGCAGATGT <u>CAT</u> CTCAACGTAAGGGTAGTC
C487R	F: GACTACCCTTACGTTGAG <u>CGT</u> ACATCTGCAACAATGGAGG R: CCTCCATTGTTGCAGATGT <u>ACG</u> CTCAACGTAAGGGTAGTC
C487A	F: GACTACCCTTACGTTGAG <u>GCT</u> ACATCTGCAACAATGGAGG R: CCTCCATTGTTGCAGATGT <u>AGC</u> CTCAACGTAAGGGTAGTC
C565M	F: TGAATACCCTTACGTTGAGAT <u>G</u> ACATCTGCAACAATGGAG R: CTCCATTGTTGCAGATGT <u>CAT</u> CTCAACGTAAGGGTAGTCA
C565R	F: TGAATACCCTTACGTTGAG <u>CGT</u> ACATCTGCAACAATGGAG R: CTCCATTGTTGCAGATGT <u>ACG</u> CTCAACGTAAGGGTAGTCA
C565A	F: TGAATACCCTTACGTTGAG <u>GCT</u> ACATCTGCAACAATGGAG R: CTCCATTGTTGCAGATGT <u>AGC</u> CTCAACGTAAGGGTAGTCA
Y535W	F: CAATGGTGGTTTCGCATCTTGGGAATTAACAAGATCTTACCC R: GGGTAAGATCTTGTTAATT <u>CCA</u> AGATGCGAAACCACCATTG
Y535L	F: CAATGGTGGTTTCGCATCTCTA <u>G</u> AATTAACAAGATCTTACCC R: GGGTAAGATCTTGTTAATTCTA <u>G</u> AGATGCGAAACCACCATTG
Y535A	F: CAATGGTGGTTTCGCATCTGCTGAATTAACAAGATCTTACCC R: GGGTAAGATCTTGTTAATT <u>CA</u> GAGATGCGAAACCACCATTG
H260R	F: TCCAGGTAGAATGTGGTGT <u>CGT</u> TGTAGAATGGTTTATTGGCC R: GGCAAATAAACCAATTCTACA <u>ACG</u> ACACCACATTCTACCTGGA
H260W	F: TCCAGGTAGAATGTGGTGT <u>TGG</u> TGTAGAATGGTTTATTGGCC R: GGCAAATAAACCAATTCTACAC <u>CA</u> ACCACATTCTACCTGGA
H260A	F: TCCAGGTAGAATGTGGTGT <u>GCT</u> TGTAGAATGGTTTATTGGCC R: GGCAAATAAACCAATTCTACA <u>AGC</u> ACACCACATTCTACCTGGA
DC486AA	F: AGATCACGGTTGGTTAATTTCTGCTGCTCTGCTGAAGGTTTAAAGGC R: GCCTTTAAACCTTCAGCAGAG <u>AGCAGC</u> AGAAATTAACCAACCGTGATCT
TAE4890AAA	F: CACGGTTGGTTAATTTCTGACTGTGCTGCTGCTGGTTTAAAGGCTGCTTAAATGTTA R: TAACATTAAGCAGCCTTTAAACCA <u>AGCAGCAGC</u> ACAGTCAGAAATTAACCAACCGTG
QW113AA	F: GTCTTTTATTCTCTATCGCTACTTCTGATGGTAACGCTGCTTCTGATTGGGTGGT R: ACCACCAAATCAGAAGCAGCGTTACCATCAGAAAGTAGCGATAGAAGAATAAAAAAGAC
QW163AA	F: GTAGATATGTTTATAATCACGCTAACGAAGATGGTGGTGCTGGTTTGCATATTGAAGGT R: ACCTTCAATATGCAAACCAGCACCACCATTCTCGTTAGCGTGATTATAACATATCTAC
QW603AA	F: GCAAATTTCTTAGAGAATATGGCTAGAACAGACGGTTCTGCTTACGGTTGTTGGGGTGGTT R: AAACACCCCAACAACCGTAAGCAGAAACCGTCTGTTCTAGCCATATTCTTAAGAAATTTGC
DCTAE4865A	F: ATCACGGTTGGTTAATTTCTGCTGCTGCTGCTGCTGGTTTAAAGGCTGCTTAAATG R: CATTAAGCAGCCTTTAAACAGCAGCAGCAGCAGCAGAAATTAACCAACCGTGAT
Site-saturation mutagenesis 260	F: TCCAGGTAGAATGTGGTGT <u>NNK</u> TGTAGAATGGTTTATT R: GGCAAATAAACCAATTCTACAM <u>NNN</u> ACACCACATTCTACCT

The underlined oligonucleotides represent mutagenized sites.

Figure A1. Two substrate prefolding types during the initial oxidosqualene cyclization reactions.

Figure A2. Homology model of cucurbitadienol synthase constructed using the human LAS crystal structure as a template. During optimization of the model no major structural changes were observed for active site residues, and refinement consisted only on fine-tuning the coordinates of active site sidechain atoms.

Figure A3. Conformations of 2,3-oxidosqualene and C-4 cation after geometry optimization. (a,b) 2,3-oxidosqualene (c,d) C-4 cation. Yellow: carbon atom; Red: oxygen atom.

Figure A4. Cont.

Figure A4. Cont.

Figure A4. (a) ^1H NMR spectrum of compound 3. (b) ^{13}C CAPT spectrum of compound 3. (c) HSQC spectrum of compound 3. (d) HMBC spectrum of compound 3. (e) ESIMS spectrum of compound 3.

Figure A5. Cont.

Figure A5. Cont.

(e)

Figure A5. (a) ^1H NMR spectrum of compound 4. (b) ^{13}C CAPT spectrum of compound 4. (c) HSQC spectrum of compound 4. (d) HMBC spectrum of compound 4. (e) ESIMS spectrum of compound 4.

(a)

Figure A6. Cont.

(b)

(c)

Figure A6. Cont.

(d)

(e)

Figure A6. (a) ^1H NMR spectrum of compound 5. (b) ^{13}C CAPT spectrum of compound 5. (c) HSQC spectrum of compound 5. (d) HMBC spectrum of compound 5. (e) ESIMS spectrum of compound 5.

Figure A7. Cont.

Figure A7. Cont.

(e)

Figure A7. (a) ^1H NMR spectrum of compound 7. (b) ^{13}C CAPT spectrum of compound 7. (c) HSQC spectrum of compound 7. (d) HMBC spectrum of compound 7. (e) ESIMS spectrum of compound 7.

References

- Hill, R.A.; Connolly, J.D. Triterpenoids. *Nat. Prod. Rep.* **2013**, *30*, 1028–1065. [[CrossRef](#)] [[PubMed](#)]
- Nes, W.R. Role of sterols in membranes. *Lipids* **1974**, *9*, 596–612. [[CrossRef](#)] [[PubMed](#)]
- Xu, R.; Fazio, G.C.; Matsuda, S.P. On the origins of triterpenoid skeletal diversity. *Phytochemistry* **2004**, *65*, 261–291. [[CrossRef](#)] [[PubMed](#)]
- Augustin, J.M.; Kuzina, V.; Andersen, S.B.; Bak, S. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **2011**, *72*, 435–457. [[CrossRef](#)] [[PubMed](#)]
- Moses, T.; Papadopoulou, K.K.; Osbourn, A. Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.* **2014**, *49*, 439–462. [[CrossRef](#)] [[PubMed](#)]
- Kovganko, N.V.; Ananich, S.K. The Chemical Synthesis of Sterols: Latest Advances. *Chem. Nat. Compd.* **1999**, *35*, 229–259. [[CrossRef](#)]
- Yu, B.; Sun, J. Current synthesis of triterpene saponins. *Chem. Asian J.* **2010**, *4*, 642–654. [[CrossRef](#)] [[PubMed](#)]
- Abe, I.; Rohmer, M.; Prestwich, G.D. Enzymatic Cyclization of Squalene and Oxidosqualene to Sterols and Triterpenes. *Chem. Rev.* **1993**, *93*, 2189–2206. [[CrossRef](#)]
- Kolesnikova, M.D.; Xiong, Q.; Lodeiro, S.; Hua, L.; Matsuda, S.P.T. Lanosterol biosynthesis in plants. *Arch. Biochem. Biophys.* **2006**, *447*, 87–95. [[CrossRef](#)] [[PubMed](#)]
- Suzuki, M.; Xiang, T.; Ohyama, K.; Seki, H.; Saito, K.; Muranaka, T.; Hayashi, H.; Katsube, Y.; Kushiro, T.; Shibuya, M. Lanosterol Synthase in Dicotyledonous Plants. *Plant Cell Physiol.* **2006**, *47*, 565–571. [[CrossRef](#)] [[PubMed](#)]

11. Eschenmoser, A.; Ruzicka, L.; Jeger, O.; Arigoni, D. Zur Kenntnis der Triterpene. 190. Mitteilung. Eine stereochemische Interpretation der biogenetischen Isoprenregel bei den Triterpenen. *Helv. Chim. Acta* **1955**, *38*, 1890–1904. [[CrossRef](#)]
12. Wu, T.K.; Griffin, J.H. Conversion of a plant oxidosqualene-cycloartenol synthase to an oxidosqualene-lanosterol cyclase by random mutagenesis. *Biochemistry* **2002**, *41*, 8238–8244. [[CrossRef](#)] [[PubMed](#)]
13. Wu, T.K.; Liu, Y.T.; Chang, C.H.; Yu, M.T.; Wang, H.J. Site-saturated mutagenesis of histidine 234 of *Saccharomyces cerevisiae* oxidosqualene-lanosterol cyclase demonstrates dual functions in cyclization and rearrangement reactions. *J. Am. Chem. Soc.* **2006**, *128*, 6414–6419. [[CrossRef](#)] [[PubMed](#)]
14. Chang, C.H.; Chen, Y.C.; Tseng, S.W.; Liu, Y.T.; Wen, H.Y.; Li, W.H.; Huang, C.Y.; Ko, C.Y.; Wang, T.T.; Wu, T.K. The cysteine 703 to isoleucine or histidine mutation of the oxidosqualene-lanosterol cyclase from *Saccharomyces cerevisiae* generates an iridal-type triterpenoid. *Biochimie* **2012**, *94*, 2376–2381. [[CrossRef](#)] [[PubMed](#)]
15. Chang, C.H.; Wen, H.Y.; Shie, W.S.; Lu, C.T.; Li, M.E.; Liu, Y.T.; Li, W.H.; Wu, T.K. Protein engineering of oxidosqualene-lanosterol cyclase into triterpene monocyclus. *Org. Biomol. Chem.* **2013**, *11*, 4214–4219. [[CrossRef](#)] [[PubMed](#)]
16. Sato, T.; Hoshino, T. Catalytic function of the residues of phenylalanine and tyrosine conserved in squalene-hopene cyclases. *Biosci. Biotechnol. Biochem.* **2001**, *65*, 2233–2242. [[CrossRef](#)] [[PubMed](#)]
17. Hart, E.A.; Ling, H.; Darr, L.B.; Wilson, W.K.; Jihai Pang, A.; Matsuda, S.P.T. Directed Evolution To Investigate Steric Control of Enzymatic Oxidosqualene Cyclization. An Isoleucine-to-Valine Mutation in Cycloartenol Synthase Allows Lanosterol and Parkeol Biosynthesis. *J. Am. Chem. Soc.* **1999**, *121*, 9887–9888. [[CrossRef](#)]
18. Herrera, J.B.R.; And, W.K.W.; Matsuda, S.P.T. A Tyrosine-to-Threonine Mutation Converts Cycloartenol Synthase to an Oxidosqualene Cyclase that Forms Lanosterol as Its Major Product. *J. Am. Chem. Soc.* **2000**, *122*, 6765–6766. [[CrossRef](#)]
19. Meyer, M.M.; Xu, R.; Matsuda, S.P. Directed evolution to generate cycloartenol synthase mutants that produce lanosterol. *Org. Lett.* **2002**, *4*, 1395–1398. [[CrossRef](#)] [[PubMed](#)]
20. Lodeiro, S.; Schulz-Gasch, T.; Matsuda, S. Enzyme Redesign: Two Mutations Cooperate to Convert Cycloartenol Synthase into an Accurate Lanosterol Synthase. *J. Am. Chem. Soc.* **2005**, *127*, 14132–14133. [[CrossRef](#)] [[PubMed](#)]
21. Salmon, M.; Thimmappa, R.B.; Minto, R.E.; Melton, R.E.; Hughes, R.K.; O'Maille, P.E.; Hemmings, A.M.; Osbourn, A. A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E4407–E4414. [[CrossRef](#)] [[PubMed](#)]
22. Ito, R.; Masukawa, Y.; Nakada, C.; Amari, K.; Nakano, C.; Hoshino, T. β -Amyrin synthase from *Euphorbia tirucalli*. Steric bulk, not the π -electrons of Phe, at position 474 has a key role in affording the correct folding of the substrate to complete the normal polycyclization cascade. *Org. Biomol. Chem.* **2014**, *12*, 3836–3846. [[CrossRef](#)] [[PubMed](#)]
23. Dai, L.; Liu, C.; Zhu, Y.; Zhang, J.; Men, Y.; Zeng, Y.; Sun, Y. Functional Characterization of Cucurbitadienol Synthase and Triterpene Glycosyltransferase Involved in Biosynthesis of Mogrosides from *Siraitia grosvenorii*. *Plant Cell Physiol.* **2015**, *56*, 1172–1182. [[CrossRef](#)] [[PubMed](#)]
24. Zhao, H.; Tang, Q.; Mo, C.; Bai, L.; Tu, D.; Ma, X. Cloning and characterization of squalene synthase and cycloartenol synthase from *Siraitia grosvenorii*. *Acta Pharm. Sin. B* **2017**, *7*, 215–222. [[CrossRef](#)] [[PubMed](#)]
25. Shibuya, M.; Adachi, S.; Ebizuka, Y. Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis. *Tetrahedron* **2004**, *35*, 6995–7003. [[CrossRef](#)]
26. Davidovich-Rikanati, R.; Shalev, L.; Baranes, N.; Meir, A.; Itkin, M.; Cohen, S.; Zimble, K.; Portnoy, V.; Ebizuka, Y.; Shibuya, M. Recombinant yeast as a functional tool for understanding bitterness and cucurbitacin biosynthesis in watermelon (*Citrullus* spp.). *Yeast* **2015**, *32*, 103–114. [[PubMed](#)]
27. Shang, Y.; Ma, Y.; Zhou, Y.; Zhang, H.; Duan, L.; Chen, H.; Zeng, J.; Zhou, Q.; Wang, S.; Gu, W. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* **2014**, *346*, 1084–1088. [[CrossRef](#)] [[PubMed](#)]
28. Thoma, R.; Schulzgasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **2004**, *432*, 118–122. [[CrossRef](#)] [[PubMed](#)]
29. Poralla, K.; Hewelt, A.; Prestwich, G.D.; Abe, I.; Reipen, I.; Sprenger, G.A. specific amino acid repeat in squalene and oxidosqualene cyclases. *Trends Biochem. Sci.* **1994**, *19*, 157–158. [[CrossRef](#)]

30. Corey, E.J.; Cheng, H.; Baker, C.H.; Matsuda, S.P.T.; Li, D.; Song, X. Studies on the Substrate Binding Segments and Catalytic Action of Lanosterol Synthase. Affinity Labeling with Carbocations Derived from Mechanism-Based Analogs of 2,3-Oxidosqualene and Site-Directed Mutagenesis Probes. *J. Am. Chem. Soc.* **1997**, *119*, 1289–1296. [[CrossRef](#)]
31. Wu, T.K.; Liu, Y.T.; Chang, C.H. Histidine residue at position 234 of oxidosqualene-lanosterol cyclase from *saccharomyces cerevisiae* simultaneously influences cyclization, rearrangement, and deprotonation reactions. *Chem. Biochem.* **2005**, *6*, 1177–1181. [[CrossRef](#)] [[PubMed](#)]
32. Weng, J.K. The evolutionary paths towards complexity: A metabolic perspective. *New Phytol.* **2014**, *201*, 1141–1149. [[CrossRef](#)] [[PubMed](#)]
33. Banta, A.B.; Wei, J.H.; Gill, C.C.; Giner, J.L.; Welander, P.V. Synthesis of arborane triterpenols by a bacterial oxidosqualene cyclase. *Proc. Natl. Acad. Sci. USA* **2016**, *114*, 245–250. [[CrossRef](#)] [[PubMed](#)]
34. Ito, R.; Hashimoto, I.; Masukawa, Y.; Hoshino, T. Effect of cation- π interactions and steric bulk on the catalytic action of oxidosqualene cyclase: A case study of Phe728 of β -amyrin synthase from *Euphorbia tirucalli* L. *Chem. Eur. J.* **2013**, *19*, 17150–17158. [[CrossRef](#)] [[PubMed](#)]
35. Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. Triterpene Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 225–257. [[CrossRef](#)] [[PubMed](#)]
36. Wendt, K.U.; Lenhart, A.; Schulz, G.E. The structure of the membrane protein squalene-hopene cyclase at 2.0 Å resolution 1. *J. Mol. Biol.* **1999**, *286*, 175–187. [[CrossRef](#)] [[PubMed](#)]
37. Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. Mutational Studies on Triterpene Synthases: Engineering Lupeol Synthase into β -Amyrin Synthase. *J. Am. Chem. Soc.* **2000**, *122*, 6818–6824. [[CrossRef](#)]
38. Schulz-Gasch, T.; Stahl, M. Mechanistic insights into oxidosqualene cyclizations through homology modeling. *J. Comput. Chem.* **2003**, *24*, 741–753. [[CrossRef](#)] [[PubMed](#)]
39. Kelly, R.; Miller, S.M.; Lai, M.H.; Kirsch, D.R. Cloning and characterization of the 2,3-oxidosqualene cyclase-coding gene of *Candida albicans*. *Gene* **1990**, *87*, 177–183. [[CrossRef](#)]
40. Ma, Y.; Zhou, Y.; Ovchinnikov, S.; Greisen, P., Jr.; Huang, S.; Yi, S. New insights into substrate folding preference of plant OSCs. *Sci. Bull.* **2016**, *61*, 1407–1412. [[CrossRef](#)]
41. Vickers, C.E.; Bydder, S.F.; Zhou, Y.; Nielsen, L.K. Dual gene expression cassette vectors with antibiotic selection markers for engineering in *Saccharomyces cerevisiae*. *Microb. Cell Fact.* **2013**, *12*, 96. [[CrossRef](#)] [[PubMed](#)]
42. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **2015**, *19*, 1639–1662. [[CrossRef](#)]
43. Huey, R.; Morris, G.M.; Olson, A.J.; Goodsell, D.S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2010**, *28*, 1145–1152. [[CrossRef](#)] [[PubMed](#)]
44. Geuflores, F.; Noureldin, H.H.; Nielsen, M.T.; Halkier, B.A. USER fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* **2007**, *35*, e55. [[CrossRef](#)] [[PubMed](#)]