PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The triterpene cyclase protein family: A systematic analysis

Silvia Racolta, P. Benjamin Juhl, Demet Sirim, and Jürgen Pleiss*

Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

ABSTRACT

Triterpene cyclases catalyze a broad range of cyclization reactions to form polycyclic triterpenes. Triterpene cyclases that convert squalene to hopene are named squalene-hopene cyclases (SHC) and triterpene cyclases that convert oxidosqualene are named oxidosqualene cyclases (OSC). Many sequences have been published, but there is only one structure available for each of SHCs and OSCs. Although they catalyze a similar reaction, the sequence similarity between SHCs and OSCs is low. A family classification based on phylogenetic analysis revealed 20 homologous families which are grouped into two superfamilies, SHCs and OSCs. Based on this family assignment, the Triterpene Cyclase Engineering Database (TTCED) was established. It integrates available information on sequence and structure of 639 triterpene cyclases as well as on structurally and functionally relevant amino acids. Family specific multiple sequence alignments were generated to identify the functionally relevant residues. Based on sequence alignments, conserved residues in SHCs and OSCs were analyzed and compared to experimentally confirmed mutational data. Functional schematic models of the central cavities of OSCs and SHCs were derived from structure comparison and sequence conservation analysis. These models demonstrate the high similarity of the substrate binding cavity of SHCs and OSCs and the equivalences of the respective residues. The TTCED is a novel source for comprehensive information on the triterpene cyclase family, including a compilation of previously described mutational data. The schematic models present the conservation analysis in a readily available fashion and facilitate the correlation of residues to a specific function or substrate interaction.

Proteins 2012; 00:000–000.

Key words: squalene-hopene cyclases; oxidosqualene cyclases; protein family database; protein engineering; sequence-structure-function relationship.

INTRODUCTION

Terpenoid cyclases catalyze the conversion of linear polyisoprenoid substrates to polycyclic terpenes in sophisticated cyclization cascades at high regioselectivity and stereoselectivity and thus catalyze key steps in the biosynthesis of sterols and triterpenes.[1] The proposed classification of cyclases on the basis of their reaction mechanisms subdivides them into Classes I and II enzymes.[1] Class I enzymes produce an allylic carbocation species via Mg^{2+} ions facilitating the abstraction of a pyrophosphate from their substrate. Class II cyclases such as the bacterial squalene-hopene cyclases (SHC) and the eukaryotic oxidosqualene cyclases (OSC) form a carbocation species by the protonation of a double bond or an epoxy group,[2] and therefore do not depend on substrates having a pyrophosphate group. Thus, they are promising enzymes for applications in industrial biotechnology.

SHCs and OSCs, belong to the triterpene cyclases and catalyze a similar reaction, but the sequence similarity between SHCs and OSCs is low. While more than 700 sequences of SHCs and OSCs have been described[3], only two crystal structures have been published yet: the SHC of *Alicyclobacillus acidocaldarius* (*Aac*SHC)[4] and the human OSC (*Hsa*OSC)[5]. Despite the low sequence similarity between members of both protein families, they

Additional Supporting Information may be found in the online version of this article.

*Abbrevation: Aac*SHC, squalene-hopene cyclase from *Alicyclobacillus acidocaldarius*; *Ath*CAS1, Cycloartenol synthase 1 from *Arabidopsis thaliana* oxidosqualene cyclase; BLAST, Basic Local Alignment Search Tool; DWARF, Data Warehouse for Analyzing Protein Families; HMM: Hidden Markov model; DSSP, Define Secondary Structure of Proteins; *Hsa*OSC, oxidosqualene cyclase from *Homo sapiens*; OSC: Oxidosqualene cyclase; SHC, Squalene-hopene cyclase; TTCED: Triterpene Cyclase Engineering Database; *Sce*ERG7: lanosterol synthase from *Saccharomyces cerevisiae*.

Grant sponsors: Bundesministerium für Bildung und Forschung (BMBF); BASF-SE.
*Correspondence to: Jürgen Pleiss, Institute of Technical Biochemistry, Allmandring 31, 70569 Stuttgart, Germany. E-mail: Juergen.Pleiss@itb.uni-stuttgart.de
Received 7 February 2012; Revised 16 March 2012; Accepted 20 March 2012
Published online 10 April 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24089

share a common fold: two connected (α/α) barrels which enclose a large hydrophobic cavity.[6] The putative substrate access channel is located in one of the (α/α) barrels, the catalytic amino acid, a protonating aspartic acid is located in the other (α/α) barrel.[7] Several conserved aromatic amino acids are located along the substrate binding cavity to stabilize the carbocation intermediates. In addition, 16 amino acid long sequence repeats, called the QW motifs, with the consensus sequence [K/R][G/A]X_{2-3}[F/Y/W][L/I/V]X_3QX_{2-5}-GXW were described to occur up to six times in SHCs and up to four times in OSCs.[8,9] The residues forming the QW motifs are located at the surface of the enzymes and have been proposed to enhance stability of the protein to cope with the highly exothermic cyclization reaction.[4]

Bacterial SHCs accept terpens of varying length from C15 to C30 as substrates.[10] The SHC from *Alicyclobacillus acidocaldarius* (*Aac*SHC) is an effective hopanoid producer under high temperature and was the first SHC gene characterized.[11] Since then, further SHC genes have been identified in other bacteria such as *Alicyclobacillus acidoterrestris*,[12] *Bradyrhizobium japonicum*,[13] *Methylococcus capsulatus*,[14] and *Zymomonas mobilis*.[13,15] In contrast to the double bond protonating SHCs, OSCs accept substrates such as oxidosqualene that contain an epoxy group. OSCs are almost exclusively produced by animals, plants, and fungi, and are further subdivided according to their products. OSCs that produce lanosterol (lanosterol synthases) are mainly produced by animals and fungi. The best characterized enzymes are the human lanosterol synthase (*Hsa*OSC)[5] and the lanosterol synthase from *Saccharomyces cerevisiae* (*Sce*ERG7).[16] Other OSCs that have been described are for example the cycloartenol synthase from *Arabidopsis thaliana* (*Ath*CAS1) and the β-amyrin synthase from *Nigella sativa*.[17] Up to now, almost 100 cyclization products have been described for OSCs.[18] This diversity makes OSCs an interesting subject for functional studies. In the pharmaceutical industry, OSCs are widely applied as targets for cholesterol-lowering drugs or antifungal agents.[19]

All OSCs and SHCs have the same catalytic residue, a protonating aspartic acid, which is embedded into two different sequence motifs: DxDD[20] and DCTAEA[21,22] in SHCs and OSCs respectively. In SHCs, the acidity of the aspartic acid is enhanced by a hydrogen bond to an adjacent histidine.[1] In OSCs, the acidity is enhanced by hydrogen bonding to two adjacent cysteines.[6] Several conserved residues in the substrate binding cavity were described to stabilize the cationic intermediates.[6] Mutational studies investigating some of these residues revealed a broad spectrum of altered, mainly shorter products.

It has been demonstrated previously that protein family databases provide a reliable framework for studying the relationships between sequence, structure, and function of enzyme families, such as the Lipase Engineering Database (*LED*)[23] and the Cytochrome P450 Engineering Database[24,25] (*CYPED*), and to identify functionally relevant residues or selectivity-determining positions by a comprehensive, systematic sequence analysis of all family members.[26,27] A similar online resource does not yet exist for triterpene cyclases. Thus, we established the Triterpene Cyclase Engineering Database (TTCED, http://www.ttced.uni-stuttgart.de) as such a resource, following the same strategy to assemble, classify, and analyze the family of triterpene cyclases.

Furthermore, to meet the rising demand for engineered biocatalysts in industrial cyclization reaction, we constructed schematic models of the SHC and OSC substrate binding cavities to facilitate protein engineering of triterpene cyclases by correlating conservational data of amino acids with a specific function or substrate interaction.

METHODS

Database setup

The Triterpene Cyclase Engineering Database (TTCED) was established within the DWARF system, a local data warehouse which was designed to integrate data on sequences, structures, and functional annotation with tools for extraction, transformation, and loading data from publicly available databases.[28] An initial sequence set was extended based on a BLAST[29] search for homologous proteins with an empirically determined E-value cut-off of 10^{-30} to establish a preliminary database. Sequences sharing 98% identical residues and originating from the same source organism were assigned to one single protein entry, and the longest, nonputative sequence was considered as reference sequence. For proteins with known structure, all PDB[30] entries were retrieved and stored. To improve consistency and quality of the classification, multiple sequence alignments were generated using CLUSTAL W.[31] False positive sequences were removed. The protein entries were classified in superfamilies and homologous families based on multiple sequence alignments, and a phylogenetic tree calculated by using PROML from the PHYLIP[32] package. The Jones-Taylor-Thornton-Model[33] was used applying default parameter settings. A further BLAST search with an *E*-value of 10^{-50} was performed and manually validated for each superfamily and homologous family. Annotation information was manually enriched by additional information on active site residues and QW-motifs.

Web interface

An online accessible version of the TTCED was generated (http://www.ttced.uni-stuttgart.de) and can be browsed by family or source organism. For each superfamily and homologous family, precalculated multisequence alignments are provided where functionally relevant amino acids are highlighted and the annotation information is displayed upon moving the cursor over the respective amino acid. Secondary structure information

was calculated using DSSP[34] and displayed within the multiple sequence alignments. For each alignment column a conservation score was calculated using PLOTCON.[35] For each family, a phylogenetic tree can be visualized by an in-house developed visualization tool and coloured either by source organism, by temperature optimum (as annotated in GenBank), by sequence length, or by other properties. All entries in the alignments and trees are linked to their respective source database entries. Family-specific HMM-profiles were calculated by HMMER (http://hmmer.janelia.org/). An archive, containing all data is provided for download. The database is permanently curated and regularly updated.

Structural analysis of *A. acidocaldarius* SHC und *H. sapiens* OSC

PDB structures of *A. acidocaldarius* SHC (*Aac*SHC) and *H. sapiens* OSC (*Hsa*OSC) in complex with the inhibitor 2-azasqualene (PDB entry 1UMP[36]) and the product lanosterol (PDB entry 1W6K[5]), respectively, were used to analyse the binding pockets. For visualization and analysis PyMOL[37] was used. Substrate-interacting residues, all of them in direct contact with the ligand, were determined using the LPC analysis software.[38] Amino acids at a distance of 5 Å or less from a substrate-interacting residue were regarded as second sphere residues.

Database analysis

For the analysis of conserved residues within the TTCED entries, three multisequence alignments were generated: separate alignments for the SHC superfamily (325 proteins) and the OSC superfamily (314 proteins), and an alignment including the SHC and the OSC superfamily (639 proteins). To enhance the alignment quality, 8 SHC sequences of homologous family SHC_1 (TTCED sequence id: 1, 2, 3, 5, 6, 911, and 912) were excluded because of a C-terminal loop region that disrupted the multiple sequence alignment. Additionally, 7 OSC sequences causing misalignments (TTCED sequence id: 308, 371, 372, 373, 429, 566, and 832) were excluded. Positions were considered as conserved if one amino acid at the respective position was found in more than 70% of the sequences in the multiple sequence alignment.

RESULTS

Phylogeny

The analysis of the multisequence alignment and the phylogenetic tree of all triterpene cyclases, resulted in two superfamilies: 325 proteins were assigned to the SHC superfamily, 314 proteins to the OSC superfamily. Within the two superfamilies, the proteins were assigned to one of 10 homologous SHC families or 10 homologous OSC families. The SHC superfamily contains mainly proteins of bacterial origin (Fig. 1). The homologous family SHC_1 (8 proteins) consists of predicted proteins from different families. The homologous family SHC_2 (29 proteins) contains the well characterized SHCs from the *Bradyrhizobium* strains BTAi1 and ORS278[39] and *Zymomonas mobilis.*[40] The biochemically and structurally well-characterized SHC from *Alicyclobacillus acidocaldarius* (*Aac*SHC) belongs to the homologous family SHC_3 (40 proteins) as well as uncharacterized SHCs from *Frankia* and *Streptomyces* species. Homologous family SHC_4 (42) consists mainly of SHCs from Geobacter species and uncultured organism. Homologous family SHC_5 (16 proteins) is the only homologous family in the SHC superfamily that contains fungal SHCs, prevalently from *Aspergillus* and *Ajellomyces* species. Characterized plant SHCs like the hydroxyhopane synthase from *Adiantum capillus-veneris*[41] and the dammaradiene synthase from *Dryopteris crassirhizoma*[42] can be found in homologous family SHC_6 (24 proteins) beside a majority of bacterial SHCs. The homologous families SHC_7 (19 proteins) and SHC_8 (13 proteins) contain predicted SHCs from different bacterial families and uncultured organism. Homologous family SHC_9 (84 proteins) is the largest family that contains SHCs from Proteobacteria. The homologous family SHC_10 consists mainly of SHCs from *Bacillus* strains.

The OSC superfamily consists mainly of proteins from metazoa, fungi, single-cell eukaryota, and plants (Fig. 1). Anyway, the homologous families OSC_1 (3 proteins) and OSC_2 (2 proteins) contain predicted OSCs from Bacteria. OSCs from Euglenozoa like the characterized lanosterol synthase from *Trypanosoma brucei*[43] form the homologous family OSC_3 (5 proteins). Homologous family OSC_4 (77 proteins) contains the characterized fungal lanosterol synthases from *Cephalosporium caerulens,*[44] *Pneumocystis carinii,*[45] and *Saccharomyces cerevisiae*[46,47] (*Sce*ERG7), and the clavaric acid producing cyclase of *Hypholoma sublateritium,*[48] as well as from other uncharacterized fungal OSCs. Homologous family OSC_5 (22 proteins) consists mainly of metazoan lanosterol synthases, originating for example from *Rattus norvegicus*[49] and from *Homo sapiens*[5] (*Hsa*OSC) which is also the only OSC with structure information. Predicted OSCs from Stramenopiles form the homologous family OSC_6. The homologous families OSC_7 to OSC_10 contain mainly OSCs from plant origin.

Among these, the homologous family OSC_7 (5 proteins) contains putative proteins from *Micromonas* and *Ostreococcus* species. The characterized cycloartenol synthases from *Adiantum capillus-veneris*[41], *Arabidopsis thaliana*[50] (*Ath*CAS1), *Betula platyphylla,*[51] and *Ricinus communis*[52] can be found in homologous family OSC_8 (58 proteins). The lanosterol synthase from *Arabidopsis thaliana*[53] and *Lotus japonicus*[54] were assigned to homologous family OSC_9 (39 proteins). Homologous family OSC_10 (100 proteins) is the largest family,

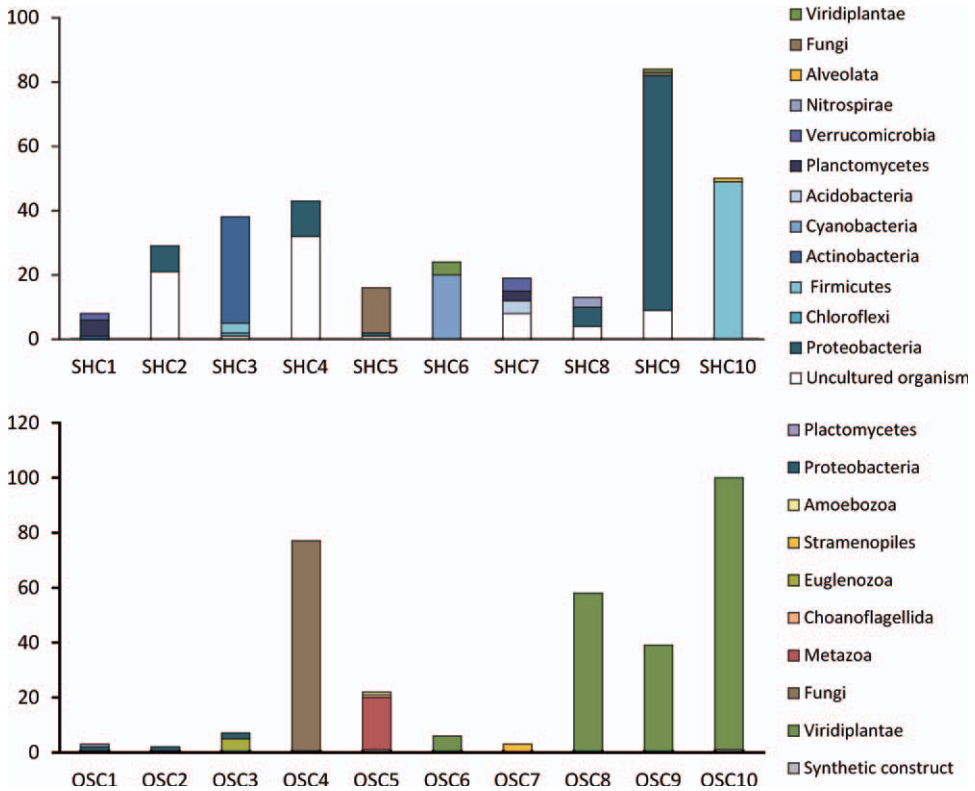

Figure 1

Taxonomic distribution in homologous families of the TTCED. The TTCED could be separated into two superfamilies: the S_SHC (upper diagram) and the S_OSC (lower diagram). The taxonomic spreading and number of members of each family is indicated by colour and size of each bar.

containing β-amyrin, cycloartenol and lupeol synthases. Characterized members of this homologous family are the β-amyrin synthase of *Nigella sativa*[17], the arabidiol,[55] and the lupeol synthase[56] of *Arabidopsis thaliana* as well as the β-amyrin synthase of *Betula platyphylla*.[51]

Substrate-interacting residues

For each superfamily and homologous family an annotated multisequence alignment was generated. Annotations were derived from the GenBank entries and manually enriched by information from literature on protonation site, carbocation stabilizing residues, and QW-motifs.

An analysis of enzyme-substrate contacts in the crystal structure of *Aac*SHC in complex with 2-azasqualene revealed 36 residues in direct contact with the substrate analogue. The analysis of *Hsa*OSC in complex with lanosterol resulted in 31 substrate-interacting residues. The amino acid conservation at these positions was analyzed by two multisequence alignments (Supporting Information Table SI). nine positions were conserved in more than 70% of all triterpene cyclases: three tryptophans, three tyrosines, two phenylalanines, and the catalytic aspartic acid (W169, W312, F365, D376, W489, Y495, F601, Y609, Y612 according to *Aac*SHC numbering; W230, W387, F444, D455, W581, Y587, F696, Y704, Y707 according to *Hsa*OSC numbering) (Table I). 12 additional substrate-interacting residues were identified that are conserved (≥70%) in the SHC superfamily (Table II). Four of them (A419, Y420, G600, and F605 according to *Aac*SHC numbering) are conserved in more than 90%. In the OSC superfamily, 13 additional positions are conserved in more than 70 % (Table III). Four of them were conserved in more than 90% of the sequences (Y98, F103, W192, and E532 according to *Hsa*OSC numbering).

The second sphere

Based on the substrate-interacting residues, a "second sphere" was defined as all residues in a distance of 5 Å or less from the substrate-interacting residues. The amino acid frequency of these positions was analyzed (Supporting Information Table SII). Totally, 131 second sphere residues were found for SHCs and 117 for OSCs, but only 9 of these positions were conserved more than 80% of all triterpene cyclases: an aspartic acid, a glycine, two prolines, two threonines, a tyrosine, and two tryptophans (A176, D313, F363, T378, T449, G487, W558, Y612, F617 according to *Aac*SHC numbering; P239, D388, F442, T457, T534, G579, W654, Y707, P712 according to *Hsa*OSC numbering). Another five positions

Table I
Residues That Are Predicted to Interact With the Substrate Analogue 2-Azasqualene (SHC) or the Product Lanosterol (OSC) and Are Conserved in All Triterpene Cyclases

Amino acid and position in *Aac*SHC[a]	Amino acid and position in *Hsa*OSC[b]	Amino acid and position in *Sce*ERG7	Amino acid and position in *Ath*CAS1	Frequency in the SHC superfamily	Frequency in the OSC superfamily
W169	W230	W232	W255	79% W; 13% S	86% W
W312	W387	W390	W416	99% W	100% W
F365	F444	F445	F472	97% F	85% F; 13% L
D376	**D455**	**D456**	**D483**	99% D	100% D
W489	W581	W587	W610	100% W	99% W
Y495	Y587	Y593	Y616	99% Y	98% Y
F601	F698	F699	F726	77% F; 16% L	89% F
Y609	Y704	Y707	Y734	97% Y	95% Y
Y612	Y707	Y710	Y737	99% Y	92% Y

[a]Structure available: PDB entry 1UMP.
[b]Structure available: PDB entry 1W6K.
Squalene-hopene cyclase from *Alicyclobacillus acidocaldarius* (*Aac*SHC) and oxidosqualene cyclases from *Homo sapiens* (*Hsa*OSC), *Saccharomyces cerevisiae* (*Sce*ERG7) and *Arabidopsis thaliana* (*Ath*CAS1) are compared. With the exception of the catalytic D (bold), all residues stabilize the cationic intermediate.

were conserved in 70% of all triterpene cyclases (P146, P429, P433, Y540, F616 according to *Aac*SHC numbering; T210, L512, P517, F633, F711 according to *Hsa*OSC numbering) (Table IV). Additional, 48 and 58 positions were conserved in the SHC superfamily, respectively in the OSC superfamily (Supporting Information Table SII).

DISCUSSION

Database size and classification

Sequence and structure information on 639 triterpene cyclases was assembled and integrated into a consistent, nonredundant database (*Triterpene Cyclase Engineering Database, TTCED*). Two superfamilies and 20 homologous families resulted from phylogenetic clustering, and family-specific annotated multisequence alignments are provided for further examinations. In a systematic search for triterpene cyclases, Frickey and Kannenberg[3] identified 729 sequences, while our analysis identified 817 sequences. The additional sequences were attributed to recent GenBank and Protein Data Base entries. Proteins differing only in few positions were grouped into a single TTCED entry, resulting in 639 unique protein entries.

Table II
Residues for the *Alicyclobacillus acidocaldarius* to Interact With the Substrate Analog 2-Azasqualene and Are Conserved More Than 70% in all SHCs

Amino acid and position in *Aac*SHC[a]	Frequency in the SHC superfamily
L36	74% L; 21% F
F129[b]	77% F; 16% M
W169[b]	79% W; 13% S
A170	74% A; 18% S
I261[b]	74% I; 22% Y
P263[b]	78% P; 11% S; 10% T
W312[b]	99% W
F365[b]	97% F
D374[b]	100% D
D376[b]	99% D
D377[b]	100% D
A419	93% A
Y420[b]	96% F
F434[b]	77% F; 12% I
W489[b]	100% W
Y495[b]	99% Y
G600[b]	94% G
F601[b]	77% F; 16% L
F605[b]	92% F
Y609[b]	97% Y
Y612[b]	99% Y

[a]Structure available: PDB entry 1W6K.
[b]Mutation published (Supporting Information Table SI).

Table III
Residues Which Are Predicted to Interact With the Product Lanosterol and Are Conserved More Than 70% in OSC

Amino acid and position in *Hsa*OSC[a]	Amino acid and position in *Sce*ERG7	Amino acid and position in *Ath*CAS1	Frequency in the OSC superfamily
Y98	Y99[b]	-	90% Y
F103	F104	F123	92% F
W192	W194[b]	W217	95% W
W230	W232[b]	W255	86% W
V236	V238	V261	77% V; 13% I
G336	A339	G366	70% G; 19% A
I338	V341	V368	79% V; 14% I
W387	W390	W416	100% W
F444	F445[b]	F472[b]	85% F, 13% L
V453	V454[b]	I481[b]	75% V; 19% I
D455	D456[b]	D483	100% D
C456	C457[b]	C484	93% C
F521	F528	F550	88% F
I524	I531	I553	77% I
E532	E539	E561	97% E
C533	C540[b]	C562	96% C
W581	W587[b]	W610	99% W
Y587	Y593	Y616	98% Y
V695	V698	V725	70% V; 12% A
F696	F699[b]	F726	89% F
Y704	Y707 [b]	Y734	95% Y

Oxidosqualene cyclases from *Homo sapiens* (*Hsa*OSC), *Saccharomyces cerevisiae* (*Sce*ERG7), and *Arabidopsis thaliana* (*Ath*CAS1) are compared.
[a]Structure available: PDB entry 1W6K.
[b]Mutation published (Supporting Information Table SI).

Table IV
Second Sphere Residues Conserved More Than 70% in SHC and OSC

Amino acid and position in *Aac*SHC[a]	Amino acid and position in *Hsa*OSC [b]	Amino acid and position in *Sce*ERG7	Amino acid and position in *Ath*CAS1	Frequency in the SHC superfamily	Frequency in the OSC superfamily
P146	T210	P212	P235	76% P; 18% F	89% P
A176	P239	P241	P264	95% P	89% P
D313	D388	D391	D417	96% D	87% D; 11 % E
F363	F442	W443	W470	94% W	94% W
T378	T457	T458	T485	83% T; 17% S	88% T
P429	L512	M519	L541	73% L	74% L
P433	P517	P524	P546	98% P	78% P; 13% A
T449	T534	T541	T563	80% T; 16% S	93% T
G487	G579	G585	G608	96% G	96% G
Y540	F633	M636	Y659	79% Y; 13% S	70% Y; 11% F
W558	W654	W657	W682	96% W	88% W
Y612	Y707	Y710	Y737	99% Y	92% Y
F616	F711	F714	F741	81% F; 13% Y	76% F; 12% Y
P617	P712	P715	P742	95% P	84% P; 11% T

Squalene-hopene cyclase from *Alicyclobacillus acidocaldarius* (*Aac*SHC) and oxidosqualene cyclases from *Homo sapiens* (*Hsa*OSC), *Saccharomyces cerevisiae* (*Sce*ERG7), and *Arabidopsis thaliana* (*Ath*CAS1) are compared.
[a]Structure available: PDB entry 1UMP.
[b]Structure available: PDB entry 1W6K.

A phylogenetic analysis demonstrated the separation into predominantly prokaryotic SHCs and eukaryotic OSCs, as pointed out by previous studies.[3] Fungal proteins were found in both superfamilies, indicating that fungi should be able to convert oxidosqualene as well as squalene. Some organisms such as *Aspergillus niger* express both SHCs and OSCs which has been supposed to be derived by lateral gene transfer.[3] Furthermore, the classification of the OSCs into homologous families according to their sequence similarity coincides well with their taxonomic classification and their product preference. In contrast, in the SHC superfamily such a strong correlation between homologous family assignment, taxonomic classification and biochemical properties cannot be found. Many SHC homologous families consist of sequences originating from different bacterial phyla. This could be due to the more common horizontal gene transfer in bacteria.

Schematic models of the substrate binding cavity

The systematic analysis of all members of the triterpene cyclase family resulted in a set of 23 residues which are highly conserved in all triterpene cyclases and therefore might have a functional or structural role. Some of these positions are known substrate-interacting residues, while other positions, mainly in the second sphere, have not been described yet. To make the information on conservation and putative function of conserved positions more readily available, we developed schematic models of the substrate binding cavity for SHCs and OSCs (Figs. 2 and 3, respectively). The model of the SHC substrate binding cavity is consistent with a model for the *Aac*SHC substrate binding cavity published previously.[57] Thus integration of the conservation data for the SHC and OSC superfamilies into the models allows easy identification of functionally relevant amino acids in SHCs and OSCs.

The reaction is initiated upon binding of the substrate to the initial substrate binding site (*Aac*SHC: W312; *Hsa*OSC: W387, T502) and by subsequent protonation by the initial protonation site (*Aac*SHC: D374, D376, Y495; *Hsa*OSC: V453, D455, Y587). The residues of the substrate binding cavity bind the substrates (*Aac*SHC: F129, I261; *Hsa*OSC: W192, G336) and stabilize the cationic intermediates (*Aac*SHC: W169, F365, D377, Y420, W489, F601, F605, Y609, Y612; *Hsa*OSC: Y98, W230, H232, T381, F444, C456, Y503, W581, F696, C700, I702, Y704, Y707). Mutants of *Aac*SHC,[20,58–71] *Sce*ERG7,[72–82] *Arabidopsis thaliana* OSC (*Ath*OSC),[77,79,83–86] *Cephalosporium caerules* OSC,[44] and *Aspergillus fumigatus* OSC[87] with amino acid exchanges at these positions have been reported to produce incomplete cyclization products with 1, 2, 3, 4, or 5 rings formed. The model represents this information by aligning the respective positions to specific areas in the model: area A for 1-ring products, area B for 2-ring products, area C for 3-ring products, area D for 4-ring products and area E for 5-ring products. For instance, the *Aac*SHC variant D377C, a mutation in area A, leads to incomplete cyclization products with 1 ring,[20] the variant F365A, a mutation in area B, leads to products with 2 rings[61] and the variant F601A, a mutation in area D, leads to products with 3 and 4 rings.[61]

The substrate binding cavity is connected to the solvent by two tunnels distal to the initial protonation site (Figs. 2 and 3). The two tunnels are visible in SHC and OSC structures (PDB entries 1UMP and 1W6K respectively). One tunnel, the distal entrance, has been suggested as a putative substrate entrance ("distal entrance"),[66] while the tunnel at the side could function as a product exit or water channel ("side exit").

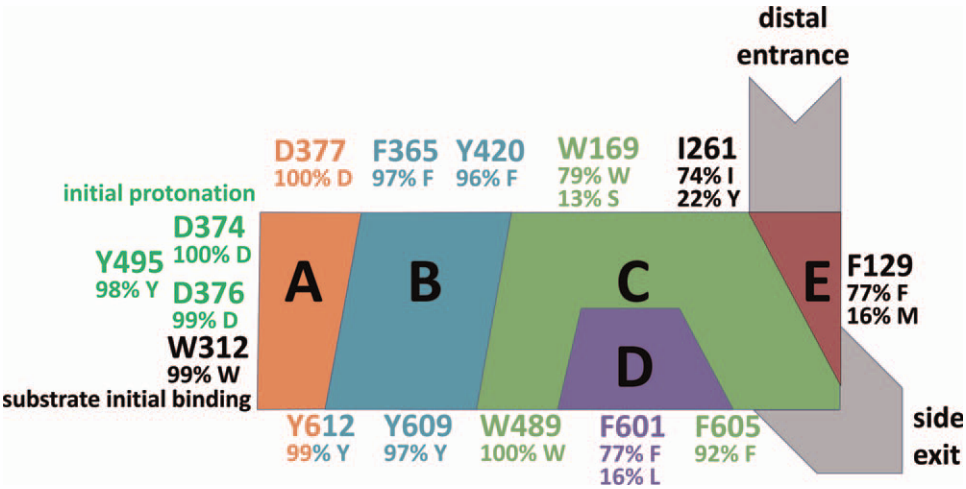

Figure 2

Schematic model of the SHC central cavity. The schematic model of the SHCs shows a putative distal substrate entrance and a possible product side exit at one end of the cavity (right side) and the residues responsible for initial protonation at the other side of the cavity (dark green, left side). Residues whose mutation leads to incomplete cyclization products are aligned to particular areas in the cavity to indicate the number of rings in these products: A, orange – 1 ring; B, blue – 2 rings; C, light green – 3 rings; D, purple – 4 rings; E, red – 5 rings. The numbering of the residues is according to the *Aac*SHC numbering.

Residues conserved in SHCs and OSCs

Despite an complete conservation of the catalytic aspartic acid (D376 and D455 in *Aac*SHC and *Hsa*OSC, respectively), the protonation sites of SHCs and OSCs differ considerably, due to the different acidity required for protonation of their substrates squalene and oxidosqualene.[4,5] In SHCs the catalytic aspartic acid is flanked by two conserved aspartic acids (D374 and D377) and is located in the DxDD motif,[1] in OSCs the catalytic aspartic acid is located within the DCTAEA motif.[6,22] Our sequence analysis demonstrated that the x in the DxDD motif predominantly is valine (47% valine, 25% isoleucine, and 14% leucine), while the second alanine in the DCTAEA motif is replaced by glycine in 52% of all OSCs. The acidity of the aspartic acid is assumed to be enhanced by a hydrogen bond to an adjacent histidine in SHCs

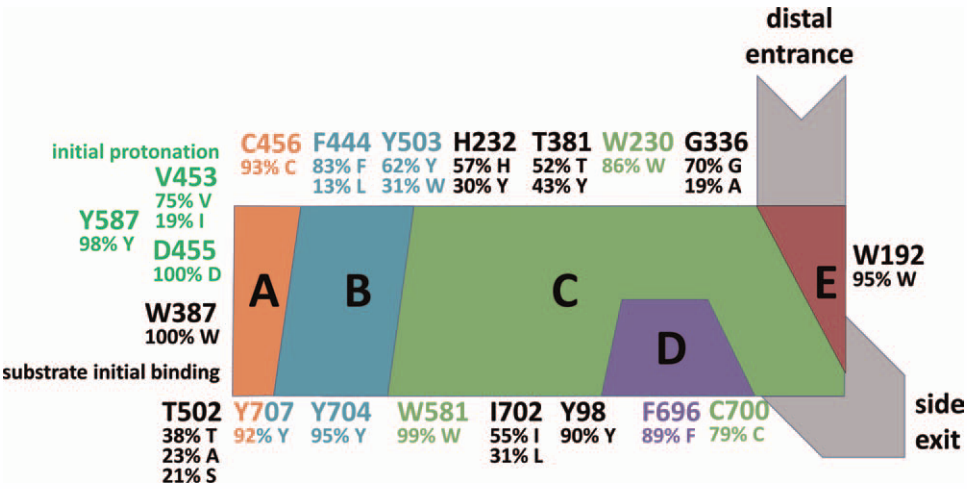

Figure 3

Schematic model of the OSC central cavity. The schematic model of the OSCs shows a putative distal substrate entrance and a possible product side exit at one end of the cavity (right side) and the residues responsible for initial protonation at the other side of the cavity (dark green, left side). Residues whose mutation leads to incomplete cyclization products are aligned to particular areas in the cavity to indicate the number of rings in these products: A, orange – 1 ring; B, blue – 2 rings; C, light green – 3 rings; D, purple – 4 rings; E, red – 5 rings. The numbering of the residues is according to the *Hsa*OSC numbering.

(H451 in *Aac*SHC)[1] or to two adjacent cysteines in OSCs (C456 and C533 in *Hsa*OSC)[6]. While both cysteines are conserved in 90% of the OSCs, the histidine is replaced by arginine in 77% of the SHCs. In accordance with this finding, the *Aac*SHC H451R variant retains 20% of its activity, while the H451F is completely inactive.[20]

Highly conserved aromatic residues in the substrate binding cavity play a crucial role in initial protonation (*Aac*SHC: Y495; *Hsa*OSC: Y587), initial substrate binding (*Aac*SHC: W312; *Hsa*OSC: W387), and cation stabilization (SHC: W169, F365, W489, F601, Y609, Y612; OSC: W230, F444, W581, F696, Y704, Y707).[16,36,57] These residues are located in the substrate binding cavity near the protonation site. Residues farther away from the protonation site show a lower conservation. This could be due to the different termination mechanisms and different product preferences in the different SHCs and OSCs.

In addition to the 9 conserved substrate-interacting residues, 14 residues in the second sphere are conserved in all triterpene cyclases. Most of the conserved second sphere residues are located near the protonation site. 4 of these second sphere residues have been analyzed by mutational studies. Mutation of the *Aac*SHC residues T378, Y540, and W558 to alanine or leucine decreased the temperature optimum of the enzyme.[20,59,60] Mutation of D313 to asparagine decreased the specific activity.[62] Thus, second sphere residues contribute to specificity, catalytic activity, or stability. However, a conservation analysis is not able to differentiate between these functions.

Residues conserved in SHCs

Apart from the residues conserved in all triterpene cyclases, additional residues are conserved only in SHCs or OSCs. Eight of the 12 substrate-interacting residues conserved in the SHC superfamily have been functionally characterized by mutagenesis studies. F129, I261, D377, Y420, and F605 (*Aac*SHC numbering) are involved in the binding, stereo-control and stabilization of the substrate.[20,58,59,63] D374 is part of the protonation site,[1] P263 and F434 are both located near the substrate entrance and have been demonstrated to play a role in termination[58,66] and in substrate gating, respectively, and are therefore not included in the schematic model of the substrate binding cavity. Position 420 mediates substrate binding and selectivity, as shown in previous mutational studies in SHC of Zymomonas mobilie[88] and in *Aac*SHC.[68] An exchange of Y420 in AacSHC by a smaller alanine or glycine led to bi- and tricyclic products highlighting the importance of this residue in positioning and stabilizing the substrate. The mutation of tyrosine to phenylalanine decreased the specific activity, while increasing the affinity to the substrate squalene.[68] Interestingly, the large majority of the SHCs (96%) have a phenylalanine at this position, such as two SHCs of *Zymomonas mobilis*.[40,89] This Gram-negative bacterium was reported to have an untypical membrane composition of bicyclic triterpenes and rearranged hopenes.[90] The presence of phenylalanine at *Aac*SHC position 420 in both *Z. mobilis* SHCs has been suggested to result in their broad product spectrum.[68]

Four additional substrate interacting residues (L36, A170, A419, and G600) have not yet been characterized by mutation studies, although A419 and G600 are conserved in more than 90% of all SHCs. Because of their high conservation and close proximity to the substrate, they might be functionally relevant and therefore are interesting positions for mutation studies.

Forty-seven second sphere positions were found to be conserved in the SHC superfamily, but only 11 positions have yet been investigated by mutation studies. H451 is important for the acidity of the protonating aspartic acid.[20] The residues W32, W406, and W417 are part of QW motifs and have an effect on the stability of the enzyme, but almost no effect on its specific activity.[60] An alanine variant of T41 showed very little differences in activity, stability and product preference, and the exact function remains unclear.[58] The Y372A and the Y606A variants both showed no effects on activity, but the Y372A variant displayed a slightly decreased temperature optimum.[59] The single mutant V381A had no effect on activity, stability, and product preference but a triple mutant containing this mutation (D277C/V380E/V381A) switch the substrate preference from squalene to oxidosqualene.[67] The three variants (D421N, D442N and E535Q) reduce the optimal temperature of *Aac*SHC.[20]

Residues conserved in OSCs

While for SHCs the formation of only a small spectrum of products has been described yet, OSCs are known to catalyze the formation of a broad range of natural products. The product spectrum has been further increased by mutations inside and outside of the substrate binding cavity. Five of the 13 conserved substrate-interacting residues have been previously investigated by mutation studies of *Saccharomyces cerevisiae* OSC *Sce*ERG7 (Y99, W194, V454, C457, and C540, corresponding to Y98, W192, V453, C456, and C533 in *Hsa*OSC). Despite the high conservation of W194 (95% W) and V454 (75% V; 19% I), the mutants W194F, V454I, and V454L showed no effect on the activity of the enzyme.[72,74,76] The role of residues Y99, C457, and C540 have been analyzed by mutation studies. A role of Y99 in the carbocation stabilization was proposed, since most of the active Y99X mutants showed incomplete cyclization products.[91] Exchanging C457 by an aspartic acid resulted in a reduced activity and it was suggested that this was mediated by modulating the acidity of the catalytic aspartic acid.[78] A similar function was suggested for C540, as an exchange to alanine did not prove crucial to the enzymes activity.[22]

However, the mutation of functionally relevant amino acids does not always result in changes of biochemical

properties. While the single mutants of V454 (corresponding to V453 in *Hsa*OSC) have no effect, the double mutants T384Y/V454I and T384Y/V454L showed a drift in the product spectrum from lanosterol to parkeol and 9α-diol.[74] Because the single mutant T384Y (corresponding to T381 in *Hsa*OSC) has no effect either,[74] the double mutants indicate a functionally relevant role of the substrate-interacting residue V454 only with the background of T384Y. At position 384 threonine and tyrosine are equally frequent (52 and 43%, respectively). Another functionally relevant position, Y510 of SceERG7 (corresponding to Y503 in *Hsa*OSC), was not found by conservation analysis, because 62% of OSCs have tyrosine, 31% tryptophan at this position. This position corresponds to the functionally relevant and highly conserved position 420 in SHCs and mutations of Y510 lead to a reduced activity and incomplete cyclization.[77] Other substrate-interacting residues that are not conserved but might be functionally relevant are H234 (57% H, 30%Y), T531 (38% T, 23% A, 21% S), and I705 (55% I, 31% L) (corresponding to H232, T502, and I702 in *Hsa*OSC). Mutation studies have confirmed that I705[81] and H234[92] play a role in the determination of the product profile. While five conserved substrate-interacting residues have been investigated by mutation studies, eight conserved residues (F103, V236, G336, I338, F521, I524, E532, and V695 according to *Hsa*OSC numbering) have not been functionally analyzed yet.

Forty-six positions in the second sphere are conserved in the OSC superfamily. One position, W198, has been investigated by mutation studies of *S. cerevisiae* lanosterol synthase *Sce*ERG7 (corresponding to W196 in *Hsa*OSC),[72] but no change in the activity or stability was found. While no second sphere residues are part of the SHC substrate binding cavity, the second sphere residue C700 is included in the OSC model, as it corresponds to the conserved substrate-interacting position F605 in SHCs. The assignment of C700 to the second sphere is probably due to the smaller size of the ligand used to determine the substrate-interacting residues in OSC. The separation between the substrate-interacting residues and the second sphere positions is not fully definite and functionally important residues can be directly interacting with the substrate or can administrate their effect in cooperation with such residues.

CONCLUSIONS

The *TTCED* represents the first integrated data source for sequence and structure information on triterpene cyclases and serves as valuable tool for the systematic and comprehensive analysis of these enzymes in the context of the whole protein family.

Especially the combination of conservation analysis and functional information provided in this work, can guide further engineering of triterpene cyclase specificity and activity.

ACKNOWLEDGMENTS

The authors thank Florian Wagner and Constantin Vogel for the implementation of the web interface and for assistance in the statistical database analysis, as well as Michael Breuer (BASF-SE) and Bernhard Hauer for valuable discussions.

REFERENCES

1. Christianson DW. Structural biology and chemistry of the terpenoid cyclases. Chem Rev 2006;106:3412–3442.
2. Wendt KU, Schulz GE. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. Structure 1998;6:127–133.
3. Frickey T, Kannenberg E. Phylogenetic analysis of the triterpene cyclase protein family in prokaryotes and eukaryotes suggests bidirectional lateral gene transfer. Environ Microbiol 2009;11:1224–1241.
4. Wendt KU, Poralla K, Schulz GE. Structure and function of a squalene cyclase. Science 1997;277:1811–1815.
5. Thoma R, Schulz-Gasch 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.
6. Abe I. Enzymatic synthesis of cyclic triterpenes. Nat Prod Rep 2007;24:1311–1331.
7. Siedenburg G, Jendrossek D. Squalene-hopene cyclases. Appl Environ Microbiol 2011;77:3905–3915.
8. Poralla K. The possible role of a repetitive amino-acid motif in evolution of triterpenoid cyclases. Bioorganic Med Chem Lett 1994;4:285–290.
9. Poralla K, Hewelt A, Prestwich GD, Abe I, Reipen I, Sprenger G. A specific amino acid repeat in squalene and oxidosqualene cyclases. Trends Biochem Sci 1994;19:157–158.
10. Hoshino T, Kumai Y, Kudo I, Nakano S, Ohashi S. Enzymatic cyclization reactions of geraniol, farnesol and geranylgeraniol, and those of truncated squalene analogs having C20 and C25 by recombinant squalene cyclase. Org Biomol Chem 2004;2:2650–2657.
11. Ochs D, Kaletta C, Entian KD, Becksickinger A, Poralla K. Cloning, expression, and sequencing of squalene-hopene cyclase, a key enzyme in triterpenoid metabolism. J Bacteriol 1992;174:298–302.
12. Wisotzkey JD, Jurtshuk P, Fox GE, Deinhard G, Poralla K. Comparative sequence analyses on the 16s ribosomal-Rna (Rdna) of *Bacillus-Acidocaldarius*, *Bacillus-Acidoterrestris*, and *Bacillus-Cycloheptanicus* and proposal for creation of a new genus, Alicyclobacillus Gen-Nov. Int J Syst Bacteriol 1992;42:263–269.
13. Perzl M, Reipen IG, Schmitz S, Poralla K, Sahm H, Sprenger GA, Kannenberg EL. Cloning of conserved genes from Zymomonas mobilis and Bradyrhizobium japonicum that function in the biosynthesis of hopanoid lipids. Biochimica Et Biophysica Acta-Lipids Lipid Metabolism 1998;1393:108–118.
14. Tippelt A, Jahnke L, Poralla K. Squalene-hopene cyclase from Methylococcus capsulatus (Bath): a bacterium producing hopanoids and steroids. Biochimica Et Biophysica Acta-Lipids Lipid Metabolism 1998;1391:223–232.
15. Reipen IG, Poralla K, Sahm H, Sprenger GA. Zymomonas-mobilis squalene-hopene cyclase gene (Shc)—cloning, DNA-sequence analysis, and expression in *Escherichia-Coli*. Microbiol-Sgm 1995;141:155–161.
16. Wu TK, Chang CH, Liu YT, Wang TT. Saccharomyces cerevisiae oxidosqualene-lanosterol cyclase: a chemistry-biology interdisciplinary study of the protein's structure-function-reaction mechanism relationships. Chem Rec 2008;8:302–325.
17. Scholz M, Lipinski M, Leupold M, Luftmann H, Harig L, Ofir R, Fischer R, Prufer D, Muller KJ. Methyl jasmonate induced accumulation of kalopanaxsaponin I in Nigella sativa. Phytochemistry 2009;70:517–522.

18. Segura MJR, Jackson BE, Matsuda SPT. Mutagenesis approaches to deduce structure-function relationships in terpene synthases. Nat Product Reports 2003;20:304–317.
19. Cravotto G, Balliano G, Robaldo B, Oliaro-Bosso S, Chimichi S, Boccalini M. Farnesyloxycoumarins, a new class of squalene-hopene cyclase inhibitors. Bioorg Med Chem Lett 2004;14:1931–1934.
20. Sato T, Hoshino T. Functional analysis of the DXDDTA motif in squalene-hopene cyclase by site-directed mutagenesis experiments: initiation site of the polycyclization reaction and stabilization site of the carbocation intermediate of the initially cyclized A-ring. Biosci Biotechnol Biochem 1999;63:2189–2198.
21. Abe I, Prestwich GD. Molecular cloning, characterization, and functional expression of rat oxidosqualene cyclase cDNA. Proc Natl Acad Sci USA 1995;92:9274–9278.
22. Corey EJ, Cheng HM, Baker CH, Matsuda SPT, Li D, Song XL. 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.
23. Fischer M, Pleiss J. The Lipase Engineering Database: a navigation and analysis tool for protein families. Nucleic Acids Res 2003;31:319–321.
24. Knoll M, Pleiss J. The medium-chain dehydrogenase/reductase engineering database: a systematic analysis of a diverse protein family to understand sequence-structure-function relationship. Protein Sci 2008;17:1689–1697.
25. Sirim D, Wagner F, Lisitsa A, Pleiss J. The Cytochrome P450 Engineering Database: integration of biochemical properties. BMC Biochem 2009;10:27.
26. Seifert A, Pleiss J. Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5. Proteins 2009;74:1028–1035.
27. Seifert A, Vomund S, Grohmann K, Kriening S, Urlacher VB, Laschat S, Pleiss J. Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. Chembiochem 2009;10:853–861.
28. Fischer M, Thai QK, Grieb M, Pleiss J. DWARF—a data warehouse system for analyzing protein families. BMC Bioinformatics 2006;7:495.
29. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
30. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58(Part 6 No 1):899–907.
31. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
32. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6: Department of Genome Sciences, University of Washington, Seattle; 2005.
33. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 1992;8:275–282.
34. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
35. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16:276–277.
36. Reinert DJ, Balliano G, Schulz GE. Conversion of squalene to the pentacarbocyclic hopene. Chem Biol 2004;11:121–126.
37. Delano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002.
38. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
39. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Vermeglio A, Medigue C, Sadowsky M. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science 2007;316:1307–1312.
40. Yang S, Pappas KM, Hauser LJ, Land ML, Chen GL, Hurst GB, Pan C, Kouvelis VN, Typas MA, Pelletier DA, Klingeman DM, Chang YJ, Samatova NF, Brown SD. Improved genome annotation for Zymomonas mobilis. Nat Biotechnol 2009;27:893–894.
41. Shinozaki J, Shibuya M, Masuda K, Ebizuka Y. Squalene cyclase and oxidosqualene cyclase from a fern. FEBS Lett 2008;582:310–318.
42. Shinozaki J, Shibuya M, Masuda K, Ebizuka Y. Dammaradiene synthase, a squalene cyclase, from Dryopteris crassirhizoma Nakai. Phytochemistry 2008;69:2559–2564.
43. Buckner FS, Nguyen LN, Joubert BM, Matsuda SP. Cloning and heterologous expression of the Trypanosoma brucei lanosterol synthase gene. Mol Biochem Parasitol 2000;110:399–403.
44. Abe I, Naito K, Takagi Y, Noguchi H. Molecular cloning, expression, and site-directed mutations of oxidosqualene cyclase from Cephalosporium caerulens. Biochim Biophys Acta 2001;1522:67–73.
45. Milla P, Viola F, Oliaro Bosso S, Rocco F, Cattel L, Joubert BM, LeClair RJ, Matsuda SP, Balliano G. Subcellular localization of oxidosqualene cyclases from Arabidopsis thaliana, Trypanosoma cruzi, and Pneumocystis carinii expressed in yeast. Lipids 2002;37:1171–1176.
46. Corey EJ, Matsuda SP, Bartel B. Molecular cloning, characterization, and overexpression of ERG7, the Saccharomyces cerevisiae gene encoding lanosterol synthase. Proc Natl Acad Sci USA 1994;91:2211–2215.
47. Shi Z, Buntel CJ, Griffin JH. Isolation and characterization of the gene encoding 2,3-oxidosqualene-lanosterol cyclase from Saccharomyces cerevisiae. Proc Natl Acad Sci USA 1994;91:7370–7374.
48. Godio RP, Martin JF. Modified oxidosqualene cyclases in the formation of bioactive secondary metabolites: biosynthesis of the antitumor clavaric acid. Fungal Genet Biol 2009;46:232–242.
49. Kusano M, Shibuya M, Sankawa U, Ebizuka Y. Molecular cloning of cDNA encoding rat 2,3-oxidosqualene: lanosterol cyclase. Biol Pharm Bull 1995;18:195–197.
50. Corey EJ, Matsuda SP, Bartel B. Isolation of an Arabidopsis thaliana gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen. Proc Natl Acad Sci USA 1993;90:11628–11632.
51. Zhang H, Shibuya M, Yokota S, Ebizuka Y. Oxidosqualene cyclases from cell suspension cultures of Betula platyphylla var. japonica: molecular evolution of oxidosqualene cyclases in higher plants. Biol Pharm Bull 2003;26:642–650.
52. Guhling O, Hobl B, Yeats T, Jetter R. Cloning and characterization of a lupeol synthase involved in the synthesis of epicuticular wax crystals on stem and hypocotyl surfaces of Ricinus communis. Arch Biochem Biophys 2006;448(1–2):60–72.
53. Kolesnikova MD, Xiong QB, Lodeiro S, Hua L, Matsuda SPT. Lanosterol biosynthesis in plants. Arch Biochem Biophys 2006;447:87–95.
54. Sawai S, Akashi T, Sakurai N, Suzuki H, Shibata D, Ayabe SI, Aoki T. Plant lanosterol synthase: Divergence of the sterol and triterpene biosynthetic pathways in eukaryotes. Plant Cell Physiol 2006;47:673–677.
55. Xiang T, Shibuya M, Katsube Y, Tsutsumi T, Otsuka M, Zhang H, Masuda K, Ebizuka Y. A new triterpene synthase from Arabidopsis thaliana produces a tricyclic triterpene with two hydroxyl groups. Org Lett 2006;8:2835–2838.
56. Herrera JB, Bartel B, Wilson WK, Matsuda SP. Cloning and characterization of the Arabidopsis thaliana lupeol synthase gene. Phytochemistry 1998;49:1905–1911.
57. Hoshino T, Sato T. Squalene-hopene cyclase: catalytic mechanism and substrate recognition. Chem Commun (Camb) 2002:291–301.

58. Sato T, Kouda M, Hoshino T. Site-directed mutagenesis experiments on the putative deprotonation site of squalene-hopene cyclase from Alicyclobacillus acidocaldarius. Biosci Biotechnol Biochem 2004;68:728–738.
59. 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.
60. Sato T, Hoshino T. Kinetic studies on the function of all the conserved tryptophans involved inside and outside the QW motifs of squalene-hopene cyclase: stabilizing effect of the protein structure against thermal denaturation. Biosci Biotechnol Biochem 1999;63:1171–1180.
61. Hoshino T, Abe T, Kouda M. Unnatural natural triterpenes produced by altering isoleucine into alanine at position 261 in hopene synthase and the importance of having the appropriate bulk size at this position for directing the stereochemical destiny during the polycyclization cascade. Chem Commun 2000:441–442.
62. Sato T, Kanai Y, Hoshino T. Overexpression of squalene-hopene cyclase by the pET vector in *Escherichia coli* and first identification of tryptophan and aspartic acid residues inside the QW motif as active sites. Biosci Biotechnol Biochem 1998;62:407–411.
63. Morikubo N, Fukuda Y, Ohtake K, Shinya N, Kiga D, Sakamoto K, Asanuma M, Hirota H, Yokoyama S, Hoshino T. Cation-pi interaction in the polyolefin cyclization cascade uncovered by incorporating unnatural amino acids into the catalytic sites of squalene cyclase. J Am Chem Soc 2006;128:13184–13194.
64. Feil C, Sussmuth R, Jung G, Poralla K. Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase. Eur J Biochem 1996;242:51–55.
65. Ceruti M, Balliano G, Rocco F, Lenhart A, Schulz GE, Castelli F, Milla P. Synthesis and biological activity of new iodoacetamide derivatives on mutants of squalene-hopene cyclase. Lipids 2005;40:729–735.
66. Milla P, Lenhart A, Grosa G, Viola F, Weihofen WA, Schulz GE, Balliano G. Thiol-modifying inhibitors for understanding squalene cyclase function. Eur J Biochem 2002;269:2108–2116.
67. Dang T, Prestwich GD. Site-directed mutagenesis of squalene-hopene cyclase: altered substrate specificity and product distribution. Chem Biol 2000;7:643–649.
68. Sato T, Sasahara S, Yamakami T, Hoshino T. Functional analyses of Tyr420 and Leu607 of Alicyclobacillus acidocaldarius squalene-hopene cyclase. Neoachillapentaene, a novel triterpene with the 1,5,6-trimethylcyclohexene moiety produced through folding of the constrained boat structure. Biosci Biotechnol Biochem 2002;66:1660–1670.
69. Full C. Bicyclic triterpenes as new main products of squalene-hopene cyclase by mutation at conserved tyrosine residues. FEBS Lett 2001;509:361–364.
70. Hoshino T, Kouda M, Abe T, Ohashi S. New cyclization mechanism for squalene: a ring-expansion step for the five-membered C-ring intermediate in hopene biosynthesis. Biosci Biotechnol Biochem 1999;63:2038–2041.
71. Full C, Poralla K. Conserved tyr residues determine functions of Alicyclobacillus acidocaldarius squalene-hopene cyclase. FEMS Microbiol Lett 2000;183:221–224.
72. Corey EJ, Matsuda SP, Baker CH, Ting AY, Cheng H. Molecular cloning of a Schizosaccharomyces pombe cDNA encoding lanosterol synthase and investigation of conserved tryptophan residues. Biochem Biophys Res Commun 1996;219:327–331.
73. Wu TK, Yu MT, Liu YT, Chang CH, Wang HJ, Diau EW. Tryptophan 232 within oxidosqualene-lanosterol cyclase from Saccharomyces cerevisiae influences rearrangement and deprotonation but not cyclization reactions. Org Lett 2006;8:1319–1322.
74. Meyer MM, Segura MJ, Wilson WK, Matsuda SP. Oxidosqualene cyclase residues that promote formation of cycloartenol, lanosterol, and parkeol. Angew Chem Int Ed Engl 2000;39:4090–4092.
75. Wu TK, Liu YT, Chiu FH, Chang CH. Phenylalanine 445 within oxidosqualene-lanosterol cyclase from Saccharomyces cerevisiae influences C-Ring cyclization and deprotonation reactions. Org Lett 2006;8:4691–4694.
76. Joubert BM, Hua L, Matsuda SP. Steric bulk at position 454 in Saccharomyces cerevisiae lanosterol synthase influences B-ring formation but not deprotonation. Org Lett 2000;2:339–341.
77. Wu TK, Chang CH. Enzymatic formation of multiple triterpenes by mutation of tyrosine 510 of the oxidosqualene-lanosterol cyclase from *Saccharomyces cerevisiae*. Chembiochem 2004;5:1712–1715.
78. Oliaro-Bosso S, Schulz-Gasch T, Balliano G, Viola F. Access of the substrate to the active site of yeast oxidosqualene cyclase: an inhibition and site-directed mutagenesis approach. Chembiochem 2005;6:2221–2228.
79. Lodeiro S, Wilson WK, Shan H, Matsuda SP. A putative precursor of isomalabaricane triterpenoids from lanosterol synthase mutants. Org Lett 2006;8:439–442.
80. Wu TK, Chang CH, Wen HY, Liu YT, Li WH, Wang TT, Shie WS. Alteration of the substrate's prefolded conformation and cyclization stereochemistry of oxidosqualene-lanosterol cyclase of Saccharomyces cerevisiae by substitution at phenylalanine 699. Org Lett 2010;12:500–503.
81. Wu TK, Chang YC, Liu YT, Chang CH, Wen HY, Li WH, Shie WS. Mutation of isoleucine 705 of the oxidosqualene-lanosterol cyclase from Saccharomyces cerevisiae affects lanosterol's C/D-ring cyclization and 17alpha/beta-exocyclic side chain stereochemistry. Org Biomol Chem 2011;9:1092–1097.
82. Wu TK, Wang TT, Chang CH, Liu YT, Shie WS. Importance of Saccharomyces cerevisiae oxidosqualene-lanosterol cyclase tyrosine 707 residue for chair-boat bicyclic ring formation and deprotonation reactions. Org Lett 2008;10:4959–4962.
83. Lodeiro S, Schulz-Gasch T, Matsuda SP. Enzyme redesign: two mutations cooperate to convert cycloartenol synthase into an accurate lanosterol synthase. J Am Chem Soc 2005;127:14132–14133.
84. Wu TK, Griffin JH. Conversion of a plant oxidosqualene-cycloartenol synthase to an oxidosqualene-lanosterol cyclase by random mutagenesis. Biochemistry 2002;41:8238–8244.
85. Meyer MM, Xu R, Matsuda SP. Directed evolution to generate cycloartenol synthase mutants that produce lanosterol. Org Lett 2002;4:1395–1398.
86. Segura MJ, Lodeiro S, Meyer MM, Patel AJ, Matsuda SP. Directed evolution experiments reveal mutations at cycloartenol synthase residue His477 that dramatically alter catalysis. Org Lett 2002;4:4459–4462.
87. Kimura M, Kushiro T, Shibuya M, Ebizuka Y, Abe I. Protostadienol synthase from Aspergillus fumigatus: functional conversion into lanosterol synthase. Biochem Biophys Res Commun 2009;391:899–902.
88. Siendenburg G, Jendroseek D, Breuer M, Juhl B, Pleiss J, Seitz M, Klebensberger J, Hauer B, Activation-independent cyclization of monoter penoids. Appl Envison Microbiol 2012;78:1055–1062.
89. Reipen IG, Poralla K, Sahm H, Sprenger GA. Zymomonas mobilis squalene-hopene cyclase gene (shc): cloning, DNA sequence analysis, and expression in *Escherichia coli*. Microbiology 1995;141(Part 1):155–161.
90. Douka E, Koukkou AI, Drainas C, Grosdemange-Billiard C, Rohmer M. Structural diversity of the triterpenic hydrocarbons from the bacterium Zymomonas mobilis: the signature of defective squalene cyclization by the squalene/hopene cyclase. Fems Microbiol Lett 2001;199:247–251.
91. Wu T-K, Li W-H, Chang C-H, Wen H-Y, Liu Y-T, Chang Y-C. Differential stereocontrolled formation of tricyclic triterpenes by mutation of tyrosine 99 of the oxidosqualene-lanosterol cyclase from *Saccharomyces cerevisiae*. Eur J Organ Chem 2009;2009:5731–5737.
92. Wu TK, Liu YT, Chang CH, Yu MT, Wang HJ. 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.