Geobiology (2007), **5**, 19–34 DOI: 10.1111/j.1472-4669.2007.00096.x

Hypotheses for the origin and early evolution of triterpenoid cyclases

W. W. FISCHER AND A. PEARSON

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

ABSTRACT

Hopanes and steranes are found almost universally in the sedimentary rock record where they often are used as proxies for aerobic organisms, metabolisms, and environments. In order to interpret ancient lipid signatures confidently we require a complementary understanding of how these modern biochemical pathways evolved since their conception. For example, generally it has been assumed that hopanoid biosynthesis was an evolutionary predecessor to steroid biosynthesis. Here we re-evaluate this assumption. Using a combined phylogenetic and biochemical perspective, we address the evolution of polycyclic triterpenoid biosynthesis and suggest several constraints on using these molecules as aerobic biomarkers. Amino acid sequence data show that the enzymes responsible for polycyclic triterpenoid biosynthesis (i.e. squalene and 2,3-oxidosqualene cyclases) are homologous. Numerous conserved domains correspond to active sites in the enzymes that are required to complete the complex cyclization reaction. From these sites we develop an evolutionary analysis of three independent characters to explain the evolution of the major classes of polycyclic triterpenoids. These characters are: (i) the number of unfavourable anti-Markovnikov ring closures, (ii) all-chair (CCC) or chair-boat-chair (CBC) substrate conformation, and (iii) the choice between squalene and 2,3-oxidosqualene as the substrate. We use these characters to construct four competing phylogenies to describe the evolution of polycyclic triterpenoid biosynthesis. The analysis suggests that malabaricanoids would be the most ancient polycyclic triterpenoids. The two most parsimonious evolutionary trees are the ones in which hopanoid and steroid cyclases diverged from a common ancestor. The transition from a CCC- to CBC-fold marks the major divergence in the evolution of these pathways, and it is diagnosable in the geological record. However, this transition does not require the simultaneous adoption of the aerobic substrate, 2,3-oxidosqualene, because these characters are controlled by independent parts of the enzyme.

Received 15 September 2006; accepted 29 November 2006

Corresponding author: W. W. Fischer. Tel.: 1-617-495-7602; fax: 1-617-495-8839; e-mail: wfischer@fas.harvard.edu.

1. INTRODUCTION

Steroids and hopanoids are two classes of polycyclic triterpenoids derived from the C_{30} isoprenoid precursor, squalene. For 50 years, chemists and biologists have been captivated by the stereoselective, enzymatic cyclization of squalene that is required to generate these products. Recently, with the discovery of a vast assemblage of lipid biomarkers in rock samples of Late Archean age [~2.7 giga-annum (Ga) to 2.5 Ga; Brocks *et al*., 1999, 2003a, b], this biosynthetic reaction has generated renewed interest among geologists and palaeontologists.

Brocks *et al.* (1999, 2003a, b) characterized a diverse collection of lipid structures, including *n*-alkanes, monomethylated alkanes, acyclic isoprenoids, cheilanthanes, and numerous cyclic triterpenoids (primarily steranes and hopanes) in sedimentary rocks from Western Australia. The presence of steranes and hopanes suggested that both eukaryotes and bacteria, respectively, were present at 2.7 Ga. In particular, the presence of a highly diverse assemblage of steranes is used as evidence for the presence of free oxygen, because their biosynthesis requires molecular oxygen in extant organisms (Tchen & Bloch, 1957; Jahnke & Klein, 1983).

Although these biomarkers appear to be syngenetic according to the evidence currently available, it remains possible that unidentified sources of contamination are the explanation for their presence in these ancient rocks (Brocks *et al.*, 2003a, c). But if the biomarkers are indeed syngenetic, their presence at 2.7 Ga is surprising. The steranes indicate the presence of at least locally oxidizing conditions at a time much earlier than a number of other redox-sensitive proxies suggest that the

atmosphere contained free O_2 (Rye & Holland, 1998; Rasmussen & Buick, 1999; Faquhar *et al.*, 2000; Bekker *et al.*, 2004; Frimmel, 2005). Thus several key questions arise: During the evolution of sterol biosynthesis, when did O_2 become a requirement? Can sterols be made without oxygen? What is the overall evolutionary history of polycyclic triterpenoids? In this paper we examine hypotheses for the evolution of the cyclization step in the triterpenoid biosynthetic pathway and evaluate in part whether these fossil biomarkers can be used to answer questions about the presence of O_2.

1.1 Are polycyclic triterpenoids biomarkers for oxygen?

The interpretation of Archean steranes and hopanes as biomarkers for oxygen rests on several assumptions. Until recently, geochemists generally accepted that hopanoids and steroids were products of aerobic bacteria and eukaryotes, respectively, both reflecting the presence of oxygen. However, evidence exists for the production of hopanoids under anaerobic conditions for organisms living in pure culture (Neunlist *et al.*, 1985; Fischer *et al.*, 2005; Härtner *et al.*, 2005) and enrichment culture (Sinninghe Damsté *et al.*, 2004). Because of uncertainty about the importance of anaerobic sources of hopanoids, the primary argument for significant O_2 in the Late Archean rests on the interpretation of sterane biomarkers.

Sterols potentially are robust biomarkers for environmental O_2 because the epoxidation of squalene requires O_2. What concentration of O_2 is needed by squalene monooxygenase (SqMO) for the production of sterols? *Methylococcus capsulatus* was able to produce 4,4-dimethyl sterols at 0.08% oxygen (Jahnke, 1986; Jahnke & Nichols, 1986), the lowest oxygen concentration tested to date. Thus, the lower limit remains unclear. However, the enzymatic reaction probably requires more O_2 than the upper limit of 10^{-5} PAL (present atmospheric level) suggested for the Archean on the basis of ancient redox proxies (Pavlov & Kasting, 2002). It appears that squalene epoxidation would require more O_2 than would have been available in the Archean atmosphere, although localized point sources (that are consumed before escaping to the atmosphere) remain a reasonable possibility.

The involvement of O_2 in modern sterol biosynthesis also is required in several downstream, post-cyclization reactions. The most abundant sterane skeletons described from the Archean are cholestanes, ergostanes, and stigmastanes (Brocks *et al.*, 2003a). Each of these products has experienced removal of three methyl groups (C_{14}, αC_4, βC_4). In extant animals, fungi, and land plants, demethylation at positions C_{14}, αC_4, and βC_4 is accomplished by O_2-requiring sterol oxidoreductases (Miller *et al.*, 1967; Alexander *et al.*, 1972). What concentration of O_2 do these enzymes require? Jahnke & Nichols (1986) partially address this question, because *M. capsulatus* is still able to demethylate position C_{14} at 0.08% oxygen, but not the methyl groups at C_4. Therefore, it appears that demethylation would also require more O_2 than would have been available in the Archean atmosphere, although again localized sources remain possible.

The apparent discordance between lipid biomarkers and other geological redox proxies led several authors to hypothesize that Archean steranes do not constrain environmental oxygen and that sterols may have been created originally *via* an unknown anaerobic pathway (Raymond & Blankenship, 2004; Kopp *et al.*, 2005). Here we create and evaluate evolutionary models to explore the evolution of sterol biosynthesis from pre-existing anaerobic reactions. The evolutionary history of different families of polycyclic triterpenoid cyclases is the primary subject of this paper. Therefore we address only the evolution of the initial O_2-requiring processes in sterol biosynthesis (epoxidation and substrate cyclization). Others have provided an analysis of downstream demethylation reactions as a palaeo-oxygen constraint (Summons *et al.*, 2006).

1.2 What are the evolutionary relationships among triterpenoid skeletons?

Because hopanoids primarily are found in bacteria and steroids primarily are found in eukaryotes, generally it is assumed that the biosynthesis of hopanoids was an evolutionary predecessor to the biosynthesis of steroids (Rohmer *et al.*, 1979, 1984; Ourisson *et al.*, 1982, 1987; Ourisson & Nakatani, 1994). Thus, many geochemists interpret the biosynthesis of steroids as *directly descendent* from hopanoid biosynthesis. While these assumptions may be valid, the biochemistry of the cyclization reaction also supports alternative scenarios. In particular, we suggest that steroids and hopanoids may have diverged separately from a common precursor, which would make their order of evolution ambiguous. The analysis also provokes an additional question: what was the ancestral polycyclic triterpenoid? The most favourable protonation of alkenes yields only tertiary carbocations (Markovnikov's rule). The compound formed when cyclization of squalene is limited to propagation only of tertiary cations is malabaricatriene, a tricyclic triterpenoid found in some anoxic environments (Behrens *et al.*, 1999; Schouten *et al.*, 2000; Werne *et al.*, 2000).

Finally, we argue that the critical divergence between the steroid and the hopanoid classes of cyclase enzymes is not defined by O_2-utilizing epoxidation, but instead by the adoption of the 'chair-boat-chair' (CBC) rather than the 'all chair' (CCC) conformation of prefolded squalene. The CBC-fold is common to all members of the steroid class as well as the pentacyclic triterpenoid, isoarborinol. Although currently there are no known cases where the two features do not co-occur, the enzymatic evidence discussed below indicates that the participation of O_2 is not directly related to the formation of this fold.

Much of the discussion presented here represents a synthesis and re-interpretation of earlier work by Rohmer *et al.* (1979,

1980, 1984), Bloch (1983), Ourisson *et al.* (1982, 1987), Ourisson & Nakatani (1994), and Kannenberg & Poralla (1999). A new analysis is warranted due to the growing availability of amino acid sequences of triterpenoid cyclases and the recent crystallization of squalene-hopene cyclase (SHC) and oxidosqualene cyclase (OSC) enzymes (Wendt *et al.*, 1997; Thoma *et al.*, 2004). The resulting evolutionary and enzymatic arguments are novel. Our hypotheses will be testable with further improvements in the record of molecular fossils, following on the work of Summons *et al.* (1988a, b), Pratt *et al.* (1991), Peters & Moldowan (1993), and Brocks *et al.* (2003a). Regardless of the final outcome of the debate about the syngeneity of Archean biomarkers, this analysis is relevant to all interpretations applied to these molecular fossils from the geological record.

2. THE CONSERVED ENZYMOLOGY OF TRITERPENOID BIOSYNTHESIS

2.1 Methods and data compilation

The biosynthesis of polycyclic triterpenoids is particularly amenable to evolutionary studies. Although the end-products produced from squalene are diverse, the pathway begins with the concerted[1] cyclization of squalene or 2,3-oxidosqualene by a triterpenoid cyclase (Woodward & Bloch, 1953; Abe *et al.*, 1993; Wendt *et al.*, 2000; Rajamani & Gao, 2003; Xu *et al.*, 2004). All known triterpenoid cyclases are remarkably homologous at the amino acid level (Perzl *et al.*, 1997, 1998; Tippelt *et al.*, 1998; Lenhart *et al.*, 2002; Bode *et al.*, 2003; Pearson *et al.*, 2003; Fischer *et al.*, 2005; this study), with expect values during similarity searches [e.g. Basic Local Alignment Search Tool (BLAST); Altschul *et al.*, 1990] typically $<10^{-70}$. There is minimal length heterogeneity between all triterpenoid cyclases, and because many of the conserved motifs represent structural domains, crystal structures of steroid and hopanoid cyclases are very similar (Wendt *et al.*, 1997; Thoma *et al.*, 2004). The high degree of enzymatic homology is due to the precise stereochemical conformation required to fold the squalene substrate and to propagate the cyclization reaction (Wendt *et al.*, 1997; Wendt & Schultz, 1998; Thoma *et al.*, 2004). The presence or absence of key residues (revealed by structural and site-directed mutagenesis studies) corresponds directly to the lipid product(s) produced by the individual enzymes.

We obtained all protein sequences of known and putative triterpenoid cyclases available from the Integrated Microbial Genomes database of the Joint Genomes Institute (JGI; http://img.jgi.doe.gov/pub/main.cgi) and from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). Parameters for BLAST were the reference SHC sequence for *Alicyclobacillus acidocaldarius* (GI: 2851526); protein query vs. translated database search (tblastn); expect value cut-off 10^{-2}; BLOSUM-62 substitution matrix; and word size 3. All species containing known and putative SHC sequences appear in Table 1. In addition, we obtained representative 2,3- OSC sequences as compiled previously (Pearson *et al.*, 2003); plus new sequences for several additional species of fungi (http://www.ncbi.nlm.nih.gov/), for the ameboid *Dictyostelium discoideum* (http://www.sanger.ac.uk/Projects/D_discoideum/), and for the diatom *Thalassiosira pseudonana* (http://genome.jgi-psf.org/diatom/). We also detected the sequence for the SHC-homolog in *Tetrahymena thermophila* that presumably is responsible for the biosynthesis of tetrahymanol, although it appears to be incorrectly annotated in the database (http://www.ciliate.org/).

Amino acid sequences were aligned using CLUSTALW (Thompson *et al.*, 1994). Multiple alignment parameters were: gap open penalty, 13.0; gap extension penalty 0.05; and BLOSUM weight matrix for proteins. The alignments were inspected, and all of the amino acids that support the key steps of the triterpenoid cyclization reaction were found to be aligned. The alignment of the regions containing conserved residues that are critical for catalysis is shown in Fig. 1 and discussed below.

2.2 Cyclization of squalene and 2,3-oxidosqualene

All of the following requirements for the synthesis of polycyclic triterpenoids are mediated directly by the cyclase enzyme. The enzyme (i) folds squalene or 2,3-oxidosqualene into the correct stereochemical configuration prior to initiating the reaction; (ii) performs the initial protonation; (iii) propagates this cation without premature quenching by the enzyme, substrate, or H_2O; (iv) creates the appropriate number of rings, often by propagating the cation against the kinetically favourable (Markovnikov) direction; and (v) quenches the cation at the final position by a Lewis base (proton acceptor) or by the addition of H_2O. The available sequence data show that this high degree of control is achieved in the same fashion by all known triterpenoid cyclases. Such homology constitutes primary evidence for the common ancestry of these enzymes and their descent from a common cyclase precursor.

The complexity of this reaction motivated the crystallization and determination of protein structure for a model SHC (*A. acidocaldarius*; Wendt *et al.*, 1997) and a model 2,3-OSC (*Homo sapiens*; Thoma *et al.*, 2004). Other squalene cyclases that yield the less-studied products tetrahymanol, isoarborinol, dammaradiene, and malabaricatriene have not been crystallized. In most cases they have not been identified by amino acid sequence, nor have they been studied *in vitro* or *in vivo*. Also, here we do not consider the more complicated

[1] There remains debate about whether the final steps of the ring closures are stepwise (e.g. Jenson & Jorgensen, 1997; Wendt *et al.*, 2000; Hoshino & Sato, 2002) or concerted (Rajamani & Gao, 2003), but this debate does not impact the present discussion.

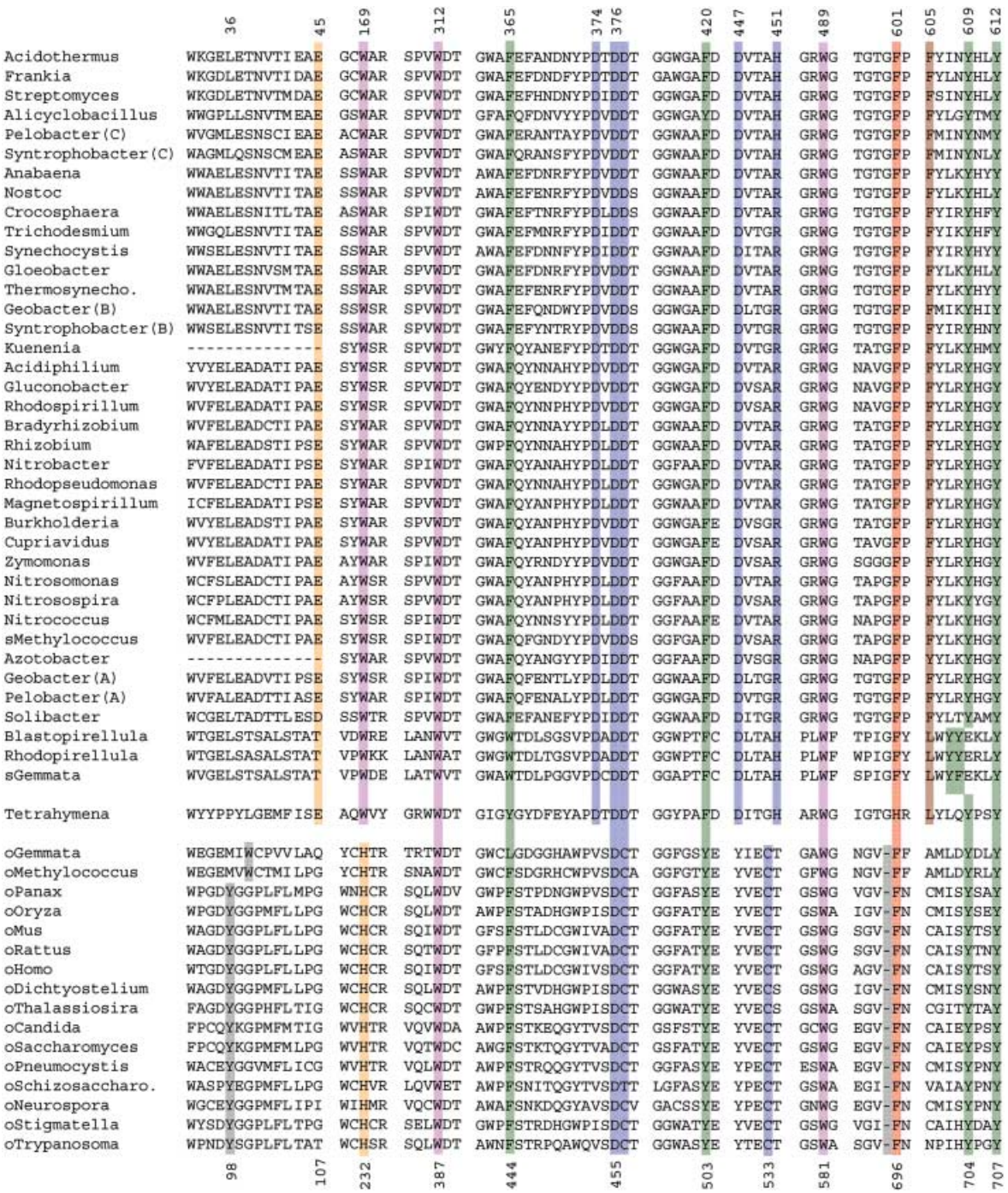

Fig. 1 Amino acid alignments for the critical functional domains of squalene-hopene cyclase (SHC), squalene-tetrahymanol cyclase (STC), and oxidosqualene cyclase (OSC) enzymes. Pink, substrate-binding residues; blue, residues for catalytic protonation and initiation of the reaction; green, residues that propagate the A-ring and B-ring carbocations; red, F601, supports the C-ring carbocation; brown, F605, supports the D-ring carbocation; grey, Tyr98, enforces boat fold of OSC; orange, His232, steroidal cation quenching. Organism information is found in Table 1. For simplicity, in most cases only one representative per genus is shown in the alignment. Prefix o before the genus name indicates the sequence is OSC.

classes of higher plant triterpenoid cyclases (Shibuya *et al.*, 1999; Xiong *et al.*, 2005), as the present discussion is concerned with the origin and evolution of cyclic triterpenoids among microbes.

2.2.1 Prefolding of squalene and 2,3-oxidosqualene

Squalene-hopene cyclase family enzymes catalyse the formation of pentacyclic hopanoids. The substrate takes on a sterically favourable all prechair (CCC) conformation (Fig. 2).

Table 1 Organisms containing triterpenoid cyclases used for the alignment in Fig. 1

GI no. (NCBI)	Cyclase	Species	Taxonomy
77128441	SHC	*Methylococcus capsulatus* BATH	γ-Proteobacteria
53611217	SHC	*Azotobacter vinelandii* AvOP	γ-Proteobacteria
77163561	SHC	*Nitrosococcus oceani* ATCC 19707	γ-Proteobacteria
82701135	SHC	*Nitrosospira multiformis* Surinam	β-Proteobacteria
30248031	SHC	*Nitrosomonas europaea* ATCC 19718	β-Proteobacteria
114330036	SHC	*Nitrosomonas eutropha* C71	β-Proteobacteria
91777110	SHC	*Burkholderia xenovorans* LB400	β-Proteobacteria
115357970	SHC	*Burkholderia ambifaria* AMMD	β-Proteobacteria
116686245	SHC	*Burkholderia cenocepacia* HI2424	β-Proteobacteria
67544393	SHC	*Burkholderia vietnamiensis* G4	β-Proteobacteria
83716035	SHC	*Burkholderia thailandensis* E264	β-Proteobacteria
77358719	SHC	*Burkholderia mallei* ATCC 23344	β-Proteobacteria
53721039	SHC	*Burkholderia pseudomallei* K96243	β-Proteobacteria
94312547	SHC	*Cupriavidus metallidurans* CH34, copy 1	β-Proteobacteria
73537298	SHC	*Cupriavidus necator* JMP134	β-Proteobacteria
94312547	SHC	*Cupriavidus metallidurans* CH34, copy 2	β-Proteobacteria
39650627	SHC	*Rhodopseudomonas palustris* CGA 009	α-Proteobacteria
82621101	SHC	*Rhizobium* sp. NGR234*	α-Proteobacteria
23011817	SHC	*Magnetospirillum magnetotacticum* MS-1	α-Proteobacteria
83309099	SHC	*Magnetospirillum magneticum* AMB-1	α-Proteobacteria
47118316	SHC	*Bradyrhizobium japonicum* USDA 110	α-Proteobacteria
75674199	SHC	*Nitrobacter winogradskyi* Nb-255	α-Proteobacteria
92115633	SHC	*Nitrobacter hamburgensis* X14	α-Proteobacteria
405607	SHC	*Zymomonas mobilis* ZM4	α-Proteobacteria
83574254	SHC	*Rhodospirillum rubrum* ATCC 11170	α-Proteobacteria
58038491	SHC	*Gluconobacter oxydans* 621H	α-Proteobacteria
88940373	SHC	*Acidiphilium cryptum* JF-5	α-Proteobacteria
39995111	SHC	†*Geobacter sulfurreducens* PCA, copy A	δ-Proteobacteria
78221228	SHC	*Geobacter metallireducens* GS-15, copy A	δ-Proteobacteria
71836816	SHC	*Pelobacter propionicus* DSM2379, copy A	δ-Proteobacteria
90960985	SHC	*Pelobacter carbinolicus* DSM 2380, copy A	δ-Proteobacteria
39995111	SHC	*Geobacter sulfurreducens* PCA, copy B	δ-Proteobacteria
78221228	SHC	*Geobacter metallireducens* GS-15, copy B	δ-Proteobacteria
88935254	SHC	*Geobacter uraniumreducens* Rf4, copy B	δ-Proteobacteria
71838408	SHC	*Pelobacter propionicus* DSM2379, copy B	δ-Proteobacteria
116747452	SHC	*Syntrophobacter fumaroxidans* MPOB, copy B	δ-Proteobacteria
90960985	SHC	*Pelobacter carbinolicus* DSM 2380, copy C	δ-Proteobacteria
116747452	SHC	*Syntrophobacter fumaroxidans* MPOB, copy C	δ-Proteobacteria
32473981	SHC‡	*Rhodopirellula baltica* SH1	Planctomycetales
87311051	SHC‡	*Blastopirellula marina* DSM3645;	Planctomycetales
50659168	SHC	*Gemmata obscuriglobus* UQM 2246;	Planctomycetales
91203347	SHC	*Candidatus 'Kuenenia stuttgartiensis'*	Planctomycetales
116222307	SHC	*Solibacter usitatus* Ellin6076, copy 1	Acidobacteria
116222307	SHC	*Solibacter usitatus* Ellin6076, copy 2	Acidobacteria
2851526	SHC	*Alicyclobacillus acidocaldarius* DSM446	Firmicutes
927383	SHC	*Alicyclobacillus acidoterrestris* DSM3922;	Firmicutes
117927211	SHC	*Acidothermus cellulolyticus* 11B;	Actinobacteria
57546753	SHC	*Streptomyces avermitilis* MA-4680;	Actinobacteria
24418971	SHC	*Streptomyces coelicolor* A3(2);	Actinobacteria
68231112	SHC	*Frankia* sp. EAN1pec, copy1;	Actinobacteria
68231572	SHC	*Frankia* sp. EAN1pec, copy 2;	Actinobacteria
86738724	SHC	*Frankia* sp. CcI3	Actinobacteria
30581838	SHC	*Nostoc punctiforme* PCC 73102	Cyanobacteria
17227497	SHC	*Anabena* sp. PCC7120	Cyanobacteria
75699950	SHC	*Anabena variabilis* ATCC 29413	Cyanobacteria
67919991	SHC	*Crocosphaera watsonii* WH8501	Cyanobacteria
110164990	SHC	*Trichodesmium erythraeum* 1MS 101	Cyanobacteria
47118304	SHC	*Synechocystis* sp. PCC 6803	Cyanobacteria
47118315	SHC	*Thermosynechococcus elongatus* BP-1	Cyanobacteria
37519569	SHC	*Gloeobacter violaceus* PCC 7421	Cyanobacteria
50659168	OSC	*Gemmata obscuriglobus* UQM 2246	Planctomycetales

Table 1 *Continued*

GI no. (NCBI)	Cyclase	Species	Taxonomy
22550400	OSC	*Methylococcus capsulatus* BATH	γ-Proteobacteria
32526539	OSC	*Stigmatella aurantiaca* DW4/3–1	δ-Proteobacteria
1352388	OSC	*Rattus norvegicus*	Metazoa
1019366	OSC	*Homo sapiens*	Metazoa
20809395	OSC	*Mus musculus*	Metazoa
6090879	OSC	*Oryza sativa*	Viridiplantae
3688602	OSC	*Panax ginseng*	Viridiplantae
28922563	OSC	*Neurospora crassa*	Fungi
170867	OSC	*Candida albicans*	Fungi
15076955	OSC	*Pneumocystis carinii*	Fungi
1169548	OSC	*Saccharomyces cerevisiae*	Fungi
1229162	OSC	*Schizosaccharomyces pombe*	Fungi
8886139	OSC	*Dictyostelium discoideum*	Mycetozoa
11023151	OSC	*Trypanosoma brucei*	Euglenozoa
15076959	OSC	*Trypanosoma cruzi*	Euglenozoa
§5.377.1	OSC	*Thalassiosira pseudonana*	Protista
n.a.	STC	*Tetrahymena thermophila*	Protista

*Encoded on a plasmid in *Rhizobium* NGR234. †δ-Proteobacteria contain three types of SHCs, designated A, B, C; each species contains two copies. ‡Presumed, but not shown, to have SHC activity. §Annotated at http://genome.jgi-psf.org/Thaps3/Thaps3.home.html newV2.0.genewise.5.377.1.

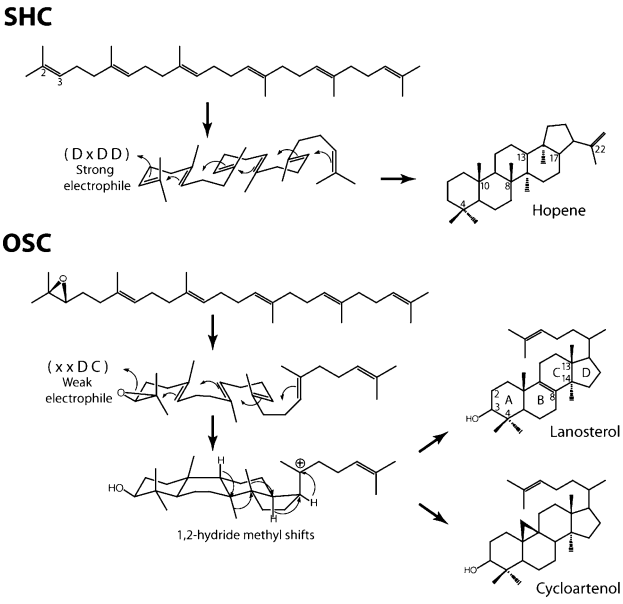

Fig. 2 Biosynthesis of polycyclic triterpenoids by squalene-hopene cyclase (SHC) and oxidosqualene cyclase (OSC). The all-chair-fold of squalene is enforced by SHC-type enzymes, while the chair-boat-chair-fold is mediated by OSC-type enzymes.

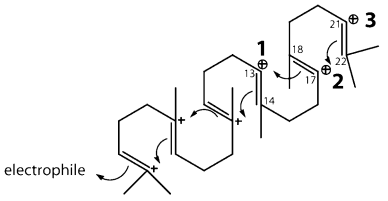

Fig. 3 Cyclization of squalene to malabaricatriene, dammaradiene, hopene, and tetrahymanol requires the promotion and stabilization of 0, 1, 2, and 3 unfavourable (2°) anti-Markovnikov carbocations, respectively.

Oxidosqualene cyclase family enzymes catalyse the formation of the initial steroidal products lanosterol, cycloartenol, and/or parkeol. The substrate adopts the sterically unfavourable CBC conformation (Fig. 3; Abe *et al.*, 1993; Wendt *et al.*, 2000).

The internal cavities of SHCs and OSCs contain numerous hydrophobic and aromatic amino acid residues to support the isoprenoid substrate (Wendt *et al.*, 1997; Thoma *et al.*, 2004). In steroid biosynthesis by OSC, the boat-fold adopted by prering B (Fig. 2) is energetically less favourable than the all-chair conformation in SHC. At least two differences in the respective amino acid sequences of SHC and OSC enzymes are responsible for enforcing these CCC- and CBC-folds, respectively (Schultz-Gasch & Stahl, 2003; Thoma *et al.*, 2004). The single residue deleted immediately before Phe696 of the *H. sapiens* sequence creates additional space beneath the mid-plane of the substrate to accommodate the –CH3 group at C_8. The –CH3 group and C_8 are forced into this boat orientation by steric hindrance created by the large Tyr98 residue above the molecular plane. Our alignment (Fig. 1) shows that these residues are conserved among all available eukaryotic OSCs, distinguishing them from SHCs. The OSC Tyr98 replaces the Lys36 present in the equivalent position of all SHC sequences (Fig. 1). Interestingly, bacterial OSCs do not contain Tyr98; instead they contain an even larger aromatic group that is offset by two AA in its location (Trp100). Presumably this residue serves the same function, as overall the

sequences align well in this region. The alternative Trp100 locus also is consistent with prior suggestions that the bacterial OSCs from *M. capsulatus* and *Gemmata obscuriglobus* are more closely related to each other than they are to eukaryotic OSCs (Pearson *et al.*, 2003).

2.2.2 Initial protonation of squalene and 2,3-oxidosqualene

The cyclization reaction initiated by SHC family enzymes begins with direct electrophilic attack and protonation of the terminal double bond of squalene. All known SHC sequences contain the strong electrophile motif: D376, assisted by D374, D377, and D447 and by the histidine/arginine (H/R) located at position 451 (DxDD motif; Fig. 1). All five of these amino acid residues have been shown by site-directed mutagenesis to be involved in the catalytic protonation of squalene (Feil *et al.*, 1996; Wendt *et al.*, 1997, 2000; Hoshino & Sato, 2002). Aspartic acid has the lowest pK_a of any amino acid, making the DxDD motif an extremely strong electrophile. This reaction by definition is anaerobic, as no external electron acceptors, metabolic cofactors, or oxygen is required.

Initiation of the reaction by OSCs is achieved by ring opening (Fig. 2), where the resonance structure of the epoxide places the oxygen atom on C_3 and the tertiary cation on C_2. All OSCs contain the weak electrophile motif x-x-Asp455-Cys456 (xxDC) in the active site (Fig. 1). When the enzyme is folded, the D455 residue of OSC is positioned in proximity to cysteines C456 and C533 (Thoma *et al.*, 2004), which aid D455 in the ring-opening reaction. The epoxide group and the extra cysteines compensate for the reduced electrophilic character of the active site. In total, five of the six amino acid residues involved in the initial protonation are different between SHC and OSC family enzymes; however, within each family all of the critical amino acids are conserved and all triterpenoid cyclases share D376/455 (Fig. 1).

Importantly, neither of these methods for creating the initial cation appears to influence the structural end-products of the reaction. The amino acids that initiate the reaction are not involved in controlling either the fold of the substrate or the propagation of the carbocation beyond the B ring. There is no evidence that they participate in forming the C, D, or E rings and that there is no imperative relationship between using 2,3-oxidosqualene as the substrate and formation of a steroidal (CBC) skeleton. SHCs also can use 2,3-oxidosqualene as a substrate, indicating that it is compatible with a CCC conformation despite its larger size. When 2,3-oxidosqualene is provided to SHC *in vitro*, the products include both 3α- and 3β-hydroxyhopanoids (Anding *et al.*, 1976; Rohmer *et al.*, 1980). Conversely, squalene presumably would be compatible with the CBC-fold of an OSC enzyme, but the reaction will not initialize because of the lack of electrophilic character in the OSC catalytic domain. Therefore, the evolution of pentacyclic and tetracyclic hydrocarbon skeletons could be independent of the involvement of oxygen in the reaction. Because the formation of the final hydrocarbon skeleton is controlled by independent parts of the enzyme, evolutionary selection did not necessarily influence the choice of substrate at the same time as it influenced the hydrocarbon structure.

2.2.3 Stabilization of A- and B-ring cations

The reaction proceeds in a zipper-like fashion, propagating the carbocation through some or all of the sites $C_4 \rightarrow C_{10} \rightarrow C_8 \rightarrow C_{13} \rightarrow C_{17}$ (hopanoid numbering; Figs 2 and 3). Throughout this chain of steps, SHC and OSC enzymes must promote and stabilize each successive cation. Accordingly, the enzymes are homologous in their conservation of numerous π-electron-rich, aromatic residues (F, W, and/or Y).

The initial carbocation created by the DxDD or xxDC active site is propagated from $C_4 \rightarrow C_{10} \rightarrow C_8$ to create the A and B rings of hopanoids and steroids (Fig. 3). Our alignment of amino acid sequences shows that formation of these two cyclohexyl rings is accomplished in an identical manner by all triterpenoid cyclases (Fig. 1). The A-ring carbocation is stabilized by Y612/707 (SHC/OSC numbering; Fig. 1), which is present in all SHCs, OSCs, and squalene-tetrahymanol cyclase (STC). The A-ring cation also is stabilized partially by the catalytic D377 (SHC numbering). Absence of D377 in SHC mutants results in products of the achilleol family (abbreviation *6*; one cyclohexane ring); and absence of D377 plus Y612 results in products of structure *6* and *6,6* (two cyclohexane rings), both due to premature quenching of the cation (Sato & Hoshino, 1999; Sato & Hoshino, 2001). These loci primarily are believed to govern the stabilization of the cation at position C_4, with a residual effect of Y612 on the second cation (position C_{10}).

The carbocation at C_{10} (B ring) also is stabilized by at least three additional conserved aromatic residues. F(Y)365/444, F(Y)420/503, and Y609/704 are found in the sequences of nearly all SHC and OSC enzymes. *Gemmata obscuriglobus* contains a lysine substitution, L444 (OSC numbering), but this is the only exception currently known (Fig. 1). It is a non-aromatic substitution, but interestingly does not impede the synthesis of sterols in this species (Pearson *et al.*, 2003). Generally, the absence of the F(Y)365/444 or Y609/704 residues results in *6,6* products (Füll & Poralla, 2000; Hoshino & Sato, 2002); the absence of the F(Y)420/503 residue results in products having *6,6* and *6,6,5* structure (Pale-Gosdemange *et al.*, 1998). These aborted cyclization products contain A and B rings, but do not proceed to form the cyclohexyl C ring.

In summary, four amino acids are critical for stabilization and closure of rings A and B (Fig. 1). These residues are common to all known SHC and OSC sequences and probably are maintained by active selection. The alignment is not affected by the differing needs of the enzyme classes to maintain the CCC- vs. CBC-folds, as both cations physically are associated with the chair fold of ring A (Fig. 2). Therefore, both SHC and OSC enzymes achieve cyclization of A and B rings through identical mechanisms, using identical amino acid motifs. A *6,6*

or *6,6,5* backbone is the product of the conserved enzymology of all polycyclic triterpenoid cyclases. Final products bearing the *6,6,5*-ring combination are related to the parent structures, 13α(H)- and 13β(H)-malabaricatriene. Functionalized malabaricanoids first were identified in the flowering plant, *Ailanthus malabarica* (Paton *et al.*, 1979), but free malabaricanes are found most commonly in association with anaerobic water columns and sediments, presumably indicating that there remain extant organisms that contain malabaricanoid cyclases (Behrens *et al.*, 1999; Schouten *et al.*, 2000; Werne *et al.*, 2000; Nytoft & Larsen, 2001; Xu *et al.*, 2004). A basic evolutionary perspective suggests that all other polycyclic triterpenoids could be considered to be derived products related to the parent structure of malabaricanoids.

2.2.4 Stabilization of C-ring cation

Steroids, hopanoids, isoarborinol, and tetrahymanol all contain a cyclohexyl C ring, which is formed by propagation of the carbocation to the first anti-Markovnikov position at C_{13} (Fig. 3). If the cyclase fails to create this secondary (2°) cation, the cation is placed in the tertiary (3°) position, yielding malabaricanoids (above). Stabilization of the anti-Markovnikov (2°) cation at position C_{13} is strongly dependent on the conserved residue F601/696 (Fig. 1). Site-directed mutagenesis experiments with the substitution F601A resulted in aborted cyclization products having the *6,6,5*-ring structure (Merkofer *et al.*, 1999; Pale-Gosdemange *et al.*, 1998). To date, all known SHCs and OSCs contain the F601 residue (Fig. 1). STC from *T. thermophila*, however, contains His601. This suggests there may be a different mechanism for ring-closure reactions in triterpenoids that need to form three anti-Markovnikov cations (rings C, D, and E). A unique mechanism for higher ring closure also is thought to operate during the synthesis of several pentacyclic plant triterpenoids (e.g. lupeols; Xiong *et al.*, 2005).

2.2.5 Stabilization of D- and E-ring cations

SHCs create a second, and in the case of STC, a third anti-Markovnikov carbocation (Fig. 3) at positions C_{17} and C_{21}, respectively, creating the cyclohexyl D and E rings. This results in the *6,6,6,6,5* and *6,6,6,6,6* structures of hopanoids and tetrahymanol/gammacerane. To support the second anti-Markovnikov cation, SHCs contain aromatic residue F605 (Fig. 1), which is not present in OSCs. Mutagenesis experiments on SHCs show that F605A produces a variety of products, including *6,6,5* and *6,6,6,5* structures (Hoshino *et al.*, 2000), consistent with the importance of this residue. The STC sequence of *Tetrahymena pyroformis* contains Lys605, rather than F605. Again a different locus must be responsible for formation of the second and eventually the third anti-Markovnikov cations to form the D and E rings of tetrahymanol.

Absence of the aromatic F605 residue in all OSC enzymes is consistent with the *6,6,6,5* backbone of steroids and the inability to create a second anti-Markovnikov cation. However, if a homologue of OSC were able to generate a second anti-Markovnikov cation, the reaction could progress to a *6,6,6,6,5* structure created from a CBC-folded substrate, e.g. isoarborinol. The amino acid residue responsible for supporting this cation would not necessarily be found in the same position (F605) as in SHC sequences, due to the differing geometry that the CBC- vs. CCC-folds would impose within the enzymatic cavity. Thus it will be difficult to diagnose a putative isoarborinol cyclase by sequence alone. Isoarborinol is a natural product derived from as-yet unknown microbial sources (Dastillung *et al.*, 1980; Ourisson *et al.*, 1982; Hauke *et al.*, 1992a, b, 1995; Jaffe & Hausmann, 1995). Future work may identify the isoarborinol cyclases and the amino acid residue(s) responsible for forming the cyclohexyl D ring of isoarborinol.

2.2.6 Cation quenching and backbone rearrangement

The terminal cation (C_{22}) formed during cyclization by SHC can be quenched by H_2O, by formation of $C\Delta^{22(29)}$ alkene, or by addition of a ribosugar at C_{29} to form bacteriohopane-polyols (Flesch & Rohmer, 1988). However, the protosterol cation is quenched differently. Following the formation of the steroidal D ring, the protosterol cation undergoes backbone rearrangement *via* concerted migration of hydride and methyl groups to yield the final products lanosterol (animals, fungi, bacteria), cycloartenol (plants, bacteria), or parkeol (bacteria). The migration of methyl groups (from sites $C_{14} \rightarrow C_{13}$ and $C_8 \rightarrow C_{14}$) is accompanied by migration of the protosterol cation from the side-chain (Fig. 2) back to C_8 where it is deprotonated by the basic residue H232 (Fig. 1; Thoma *et al.*, 2004). The role of the OSC enzyme in the rearrangement of the methyl groups attached to C_8 and C_{13} is considered to be one of geometry, as the methyl groups are induced to migrate by the steric difficulty associated with the CBC-fold. Thus, the presence of a methyl group at position C_{13} in tetracyclic *6,6,6,5* triterpenoids is diagnostic of a CBC-folded substrate. As such, a four-ring sterane (methyl groups at C_{13} and C_{14}; Fig. 2) directly reflects the CBC character; while a *6,6,6,5* dammarane-like structure with methyl groups at C_8 and C_{14} reflects an all-CCC-fold.

3. EVOLUTIONARY HYPOTHESES

Previously it has been assumed that hopanoid biosynthesis evolved early within the bacteria, sterol biosynthesis evolved later (either within the bacteria or eukaryotes), and that isoarborinol might represent an intermediate polcyclic triterpenoid structure in a linear evolutionary progression (Rohmer *et al.*, 1979; Ourisson *et al.*, 1987). The less-studied dammaranoids and malabaricanoids have been de-emphasized; however, all of these molecules potentially are significant to the evolution of triterpenoid cyclases, as all are known natural products. The homology of the amino acid alignments above suggests that triterpenoid cyclases share a common ancestry

and can be analysed within evolutionary models that are not limited to a discussion of only hopanoids and steroids.

In order to postulate the evolutionary history of triterpenoid cyclase enzymes, we constructed a theoretical framework using chemical reasoning and principles of molecular evolution (e.g. Page & Holmes, 1998; Benner, 2003). We assume that the earliest enzymes mediated the simplest and most energetically favourable reactions. We then assume – consistent with the principles of natural selection – that subsequent evolution increased the ability of the enzymes to overcome successive energetic obstacles only if the new end-products conferred some type of selective advantage to the host. The specific changes in amino acid sequence that are needed to overcome these energetic obstacles can be categorized as three independent types of character-state transitions. Below we develop and use the characters to assess the number and types of changes that would have been required for different evolutionary scenarios. There appear to be four most likely scenarios that could explain the evolution of polycyclic triterpenoids. Two, nominally, are consistent with traditional hypotheses and two are novel. All four hypotheses should be testable in the future with refined biomarker data from the geological record and with the discovery of new cyclase enzymes in extant organisms.

3.1 Defining evolutionary characters

Figure 4 shows the relationships between the major classes of polycyclic triterpenoid hydrocarbon skeletons. Important changes in the evolution of polycyclic triterpenoids are: (i) differences in the number of unfavourable anti-Markovnikov ring closures (Fig. 4; events A, A′, A″, -A‴), (ii) the change to a CBC conformation from a CCC conformation of the preformed substrate (Fig. 4, events B and B′), and (iii) the substitution of 2,3-oxidosqualene for squalene as the substrate (Fig. 4, event C). Each is chemically different and can be called a change in *character state*, because these individual processes are controlled by different amino acids. Each character therefore represents an independent degree of freedom in evolutionary scenarios.

Many combinations of character state transitions hypothetically could explain the evolution of polycyclic triterpenoids. For example, the evolution of steroids could follow the path A → B → C, or alternatively A → A′ → B′ → -A‴ → C (Fig. 4). The versions A → C → B and A → A′ → C → B′ → -A‴ also are possible if we assume that the change in substrate to 2,3-oxidosqualene preceded the evolution of the CBC-fold. In the paths described above, it is assumed that eventually it will be possible to distinguish whether the CBC event (B) preceded or followed the change in substrate (C). To interpret all ancient steranes as proxies for environmental oxygen, C must originate prior to or at the same time as B, and it must require molecular oxygen (section *3.3.3*). Currently, the B and C characters cannot be resolved, but there is no apparent enzymatic requirement that necessitates them evolving in concert. This ambiguity affects the interpretation of ancient biomarkers and it demands that these characters be counted as separate events. They cannot have happened through a single amino acid substitution or alteration of only a single structural motif within a cyclase.

The above scenarios can be portrayed as evolutionary trees. The topologies are based on the order of acquisition or loss of individual characters, with subsequent radiation of the cyclases from nodes defined by these events. We used both minimum evolution analysis and energetic considerations to generate four phylogenetic trees (Fig. 5). In all cases, malabaricanoid cyclase is placed as the most ancient triterpenoid cyclase (as argued in section *2.2.3*). In the higher branches, two of the trees are consistent with the idea that steroid biosynthesis is derived from hopanoid biosynthesis (Fig. 5C,D). However, two additional topologies imply that both steroid and hopanoid cyclases diverged separately from a common ancestor (Fig. 5A,B).

3.2 Minimum evolution models

We first presumed that the original state (root) was linear squalene; i.e. that the substrate existed prior to the evolution of any cyclase. The alternative is that a proto-cyclase existed but was specific for an alternate substrate such as hexaprene,

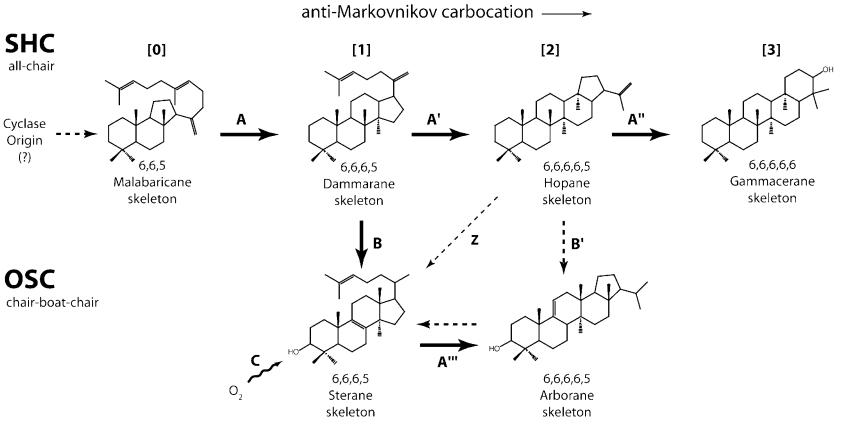

Fig. 4 Major categories of polycyclic triterpenoid hydrocarbon skeletons arranged according to the three fundamental evolutionary characters: the ability to promote and stabilize unfavourable anti-Markovnikov carbocations (A, A′, A″, A‴); enforcement of CBC backbone stereochemistry (B, B′); and the evolution of an oxygen-containing substrate (C). These relationships were used to create the evolutionary trees shown in Fig. 5.

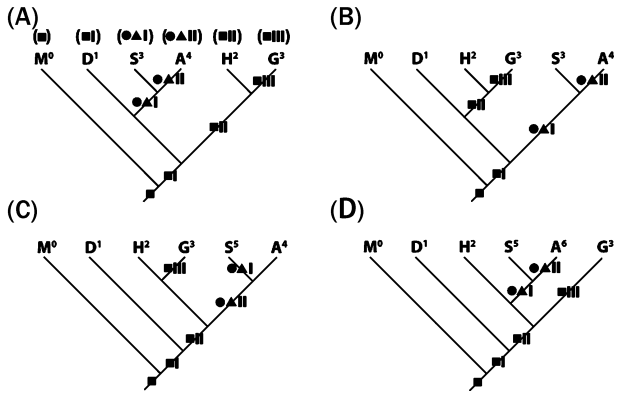

Fig. 5 Four hypothetical scenarios for the evolution of the major classes of polycyclic triterpenoid cyclases. Character symbols: square, CCC-fold; triangle, CBC-fold; circle, 2,3-oxidosqualene substrate; bar(s), anti-Markovnikov carbocation(s). *M*, malabaricatriene cyclase; D, dammaradiene cyclase; H, hopanoid cyclase; G, gammacerane/tetrahymanol cyclase; S, steroid cyclase; A, isoarborinol cyclase.

thus yielding compounds such as cheilanthanes. Regardless, the first character adopted is to prefold the substrate into either CCC or CBC conformation, as the reaction cannot proceed without a folded substrate. Based on energetic favourability, we assume that the initial character state was CCC (square symbol, Fig. 5). We also assume that squalene, not 2,3-oxidosqualene, was the initial substrate. This implies that substitution of 2,3-oxidosqualene is a derived character (circle symbol, Fig. 5). We justify the latter assumption in section 3.3. These initial conditions imply that the first polycyclic triterpenoid product was a *6,6* or a *6,6,5* compound such as malabaricatriene, produced by a malabaricatriene cyclase (M). These natural products only require the enzyme to fold (CCC), initiate protonation of squalene (DxDD motif), and support the A- and B-ring cations (section *2.2.3*).

The subsequent evolution of the dammaranoid, hopanoid, steroid, isoarborinol, and tetrahymanol cyclases requires a series of changes in the defined enzymatic characters. These changes are achieved by substituting the amino acids critical for each subsequent step of the cyclization reaction (Fig. 1). In all four scenarios shown in Fig. 5, the first new character would have been the ability to propagate one anti-Markovnikov cation (bar symbol). This character is common to all polycyclic triterpenoid cyclases other than M, and it was acquired through a single amino acid substitution, F601 (SHC numbering), which is conserved among all known cyclases. However, as the CBC-fold has not yet been acquired, the propagation of one anti-Markovnikov cation would yield dammaranoids, the primary products of a dammaranoid cyclase (D).

After the acquisition of the first anti-Markovnikov carbocation, the subsequent steps differ according to the order in which successive characters are acquired. This results in four proposed options for the radiation of hopanoid cyclases (H), tetrahymanol/gammacerane cyclases (G), sterol cyclases (S), and isoarborinol cyclases (A).

Subsequent acquisition of a second anti-Markovnikov carbocation (F605 substitution) would have led to the evolution of H from D, symbolized by a second bar (Fig. 5A). Evolution of a third anti-Markovnikov cation would have caused the divergence of G from H. In this scenario, S evolved directly from D after D and H diverged. Evolution of S requires conserving the single anti-Markovnikov cation of D while adopting the CBC-fold (triangle symbol; Y98 and mid-plane deletion below the substrate) and 2,3-oxidosqualene as a substrate (circle symbol; xxDC motif). Finally, A would have evolved from S with the independent acquisition of a second anti-Markovnikov cation in the OSC family, which would be unrelated to the second anti-Markovnikov cation of H and G in the SHC family. This cation therefore may not be supported by the same F605 residue as found in SHC sequences, as it would not share a common history with the F605 substitution. Thirteen total changes in character state are required to create the tree shown in Fig. 5(A).

Figure 5(B) is similar to Fig. 5(A): the trees differ only in the order in which the second anti-Markovnikov character is acquired. Figure 5(A) presumes that both dammaranoid and hopanoid cyclases existed before a steroid cyclase subsequently emerged from D through acquisition of the CBC-fold and oxidized substrate. In contrast, the scenario in Fig. 5(B) is the version in which S emerged from D before either of the major cyclase lineages acquired a second anti-Markovnikov cation. These scenarios should be testable against the geological and/or phylogenetic records: the former predicts that hopanoids existed before steroids, while the latter predicts that steroids existed before hopanoids. In Fig. 5(B) all subsequent steps remain the same as the scenario in Fig. 5(A). Again, G was derived from H, and A was derived from S. This scenario also requires 13 changes in character state, as only the order of evolutionary events has changed with respect to Fig. 5(A).

Figure 5(A,B) represent equally parsimonious relationships, each requiring the minimum number of evolutionary steps to derive the major classes of polycyclic triterpenoids. Both hypotheses imply that H and S evolved from a common ancestor D, a dammaranoid cyclase. The different relative timing implied by the two scenarios is significant geologically, however, as Fig. 5(A) suggests that sterols evolved late, and Fig. 5(b) suggests that sterols evolved early. Figure 5(A) is consistent with a long-lived anaerobic history of polycyclic triterpenoids that predate the appearance of 2,3-oxidosqualene (circle symbol), related to the oxygenation of Earth surface environments. This predicts an early biomarker record that would contain hopanes in the absence of steranes. However, if the scenario in Fig. 5(B) accurately represents the order of evolution, then all of the polycyclic triterpenoid cyclases other than M and D would have radiated later in Earth's history, including H. This predicts a biomarker record containing steranes in the absence of hopanes, or possibly the emergence of both at approximately the same time; and it implies that both would have arisen after the evolution of oxygenic photosynthesis.

Another option to explain Fig. 5(B) would be to invoke an anaerobic pathway for the early, direct synthesis of steranes or for the formation of 2,3-oxidosqualene by peroxidation (section *3.3.3*). Regardless, the two hypotheses shown in Fig. 5(A,B) may be directly testable within the limits of the biomarker record.

Figure 5(C) describes the hypothetical scenario most similar to the one proposed by Ourisson *et al.* (1987), which suggested that S may have evolved from H via the apparent structural intermediate, A. This order of events places arborinol cyclases (A) as descendents of the H family. To evolve A from H would require conserving both anti-Markovnikov carbocations (two bar symbols) while switching to 2,3-oxidosqualene (circle symbol) and acquiring the unfavourable CBC conformation (triangle symbol). S would then evolve subsequently from A by retrograde loss of an anti-Markovnikov cation (event -A‴, Fig. 4). The tree in Fig. 5(C) is not as parsimonious as the trees in Fig. 5(A,B), requiring 15 total changes in character state. This pathway to sterols via isoarborinol is more complicated than the direct emergence of S from a D precursor. However, this parsimony argument assumes that all character state changes come at the same biochemical cost, which may not be a valid assumption. Presumably it is harder to gain an anti-Markovnikov carbocation than it is to lose one, and therefore the reversion of A to S (event -A‴, Fig. 4) may not be as difficult as the other direction from S to A (+ A‴, Fig. 4). However, there is an additional reason why the isoarborinol intermediate scenario (B′ → -A‴) is less plausible: the amino acids responsible for π-cation stabilization of the D ring are likely to no longer be in the correct position (F605) after the CBC-fold is adopted in event B′. The residue that stabilizes the cation of the isoarborinol D-ring probably is found in a location other than position 605, due to the geometry imposed by the CBC-fold. Therefore the first CBC products probably contained fewer than five rings, and if the CBC-fold was adopted at the stage of H, the resulting effect may have been to bypass isoarborinol entirely. This bypass is shown as event Z (Fig. 4).

The consequences of the Z event are shown in Fig. 5(D), which depicts the only scenario in which hopanoid biosynthesis would be the direct antecedent of steroid biosynthesis. Throughout this analysis we have assumed that the evolutionary characters always behave independently. This is supported by experimental evidence that discrete amino acids mediate each biochemical step of the cyclization reaction (section 2). However, the Z path highlights a situation where two characters could be directly linked: adopting the CBC-fold could cause a simultaneous loss of the second anti-Markovnikov cation (Fig. 4). Such a situation is consistent with the phylogeny shown in Fig. 5(D). Here, S would emerge directly from H, and then A would emerge subsequently from S. The evolutionary tree as shown in Fig. 5(D) requires 17 changes in character state. However if we assume that the adoption of a CBC-fold causes loss of an anti-Markovnikov cation, then these changes cannot be counted as separate events and the total number would drop to 15. In either instance, the trees shown in Fig. 5(C,D) remain less parsimonious than Fig. 5(A,B). However, these trees are the ones that more closely resemble traditional hypotheses about the evolution of hopanoids and steroids. Renewed efforts by the organic geochemistry and geobiology communities to resolve the early biomarker record and to find the organisms and proteins responsible for the biosynthesis of malabaricanoids, dammaranoids, and isoarborinol may help resolve these questions.

3.3 Congruence of evolutionary proposals with known biochemistry

3.3.1 Difficult control and high mutability of anti-Markovnikov cations

Anti-Markovnikov cations are difficult to induce and stabilize. Non-enzymatic systems fail to duplicate the complex cyclization of squalene (van Tamelen *et al.*, 1966; Wendt *et al.*, 2000; Rajmani & Gao, 2003). *In vivo*, squalene cyclases exert imperfect control on the cation propagation step, even when the critical amino acids are conserved. The wild type SHC from *A. acidocaldarius* occasionally fails to promote the second anti-Markovnikov cation, resulting in 1% of the natural products having an accidental *6,6,6,5* dammarane-type skeleton (Pale-Gosdemange *et al.*, 1998; Wendt *et al.*, 2000; Rajmani & Gao, 2003). Similar products were observed from *Zymomonas mobilis* (Douka *et al.*, 2001) and also were attributed to deficient control by the enzyme. Interestingly, *Bradyrhizobium japonicum* and *Rhodopseudomonas palustris* produce minor amounts of tetrahymanol in addition to hopanoids (Kleemann *et al.*, 1990; Bravo *et al.*, 2001). Because both of these bacteria only contain one triterpenoid cyclase gene (Table 1), the single enzyme must be responsible for producing both hopane and gammarcerane-type triterpenoids.

These examples demonstrate the difficulty of controlling the formation of tetracyclic and pentacyclic products, suggesting that both gain and loss of these characters are difficult to accomplish with a sufficient degree of biosynthetic precision. This suggests there would have been a significant evolutionary barrier to altering the number of anti-Markovnikov cations without sacrificing the production of a dominant single end-product. The 2° cations themselves are energetically unfavourable, and as such it is reasonable to presume an order of evolution from fewer to more anti-Markovnikov cations. These biochemical challenges also suggest that minimizing the total number of anti-Markovnikov cations that must be changed over time is evolutionarily more parsimonious. Consistent with this, the evolutionary trees in Fig. 5(A,B) do not require the enzyme to lose and then subsequently re-gain anti-Markovnikov cations as required in Fig. 5(C,D).

3.3.2 Adapting to 2,3-oxidosqualene as a substrate

Generally it is assumed that metabolisms requiring molecular oxygen are more derived, because surface environments

during the first half of Earth's history were without significant free oxygen. The biochemistry of triterpenoid biosynthesis also is consistent with a transition from the squalene (anaerobic) to 2,3-oxidosqualene (aerobic) substrate. The active site of SHC will protonate and cyclize both the 3(R) and the 3(S) enantiomers of squalene epoxide (Anding *et al.*, 1976; Rohmer *et al.*, 1980). The OSC from *M. capsulatus* (a γ-proteobacterium containing both SHC and OSC) also will cyclize both enantiomers (Rohmer *et al.*, 1980). In contrast, in animals, yeast, and plants, all known OSCs only will cyclize 3(S)-oxidosqualene and not the 3(R) enantiomer (Barton *et al.*, 1975). The enzyme in *M. capsulatus* is less substrate-specific than the eukaryotic version and may be a relic of the early evolution of OSC genes. When a proto-OSC began to encounter 2,3-oxidosqualene regularly, the strong electrophilic character may have been lost due to relaxed constraints on the amino acids present in the protonating region. This would have rendered the enzyme suitable for cyclization of squalene epoxide only, and eventually of the 3(S) isomer only. This evolutionary transition may have been advantageous to avoid confusion between the two substrates. Loss of the DxDD electrophilic character would have allowed efficient channelling of squalene and 2,3-oxidosqualene to the appropriate enzymes. It remains unknown what purpose the dual existence of hopanoids and steroids serves for bacterial species such as *M. capsulatus*, but both are produced simultaneously (Bird *et al.*, 1971a, b; Rohmer *et al.*, 1980). Interestingly, this scenario also would be consistent with a gene duplication of an original dammaranoid cyclase, followed by evolutionary selection of the two genes toward distinct SHC and OSC functions as described in both Fig. 5(A,B). It also suggests the possibility that the ancestral dammarane cyclase may no longer exist if it was lost due to gene duplication and paralogous evolution towards more physiologically suitable polycyclic triterpenoid products.

3.3.3 Synthesis of 2,3-oxidosqualene

In the sterol pathway, squalene is epoxidized by squalene monooxygenase to form 2,3-oxidosqualene (Fig. 6), using O_2 and reducing power supplied by NADPH (Tchen & Bloch, 1957; Yamamoto & Bloch, 1970). The reaction begins with the reduction of O_2 to form a hydroperoxy group on the enzymatically bound cofactor flavin adenine dinucleotide (FAD) (Fig. 6, inset 1a). The O in the terminal position is inserted across the 2(3) double bond of squalene creating the epoxide (Fig. 6, inset 1b). The residual hydroxyflavin is reduced back to FAD and H_2O using electrons from NADPH-cytochrome P450 reductase (Ono & Bloch, 1975; Fig. 6, inset 1c). There are no currently known steroid cyclases that use squalene as the substrate, and thus all known sterol producers require O_2 and the enzyme squalene monooxygenase (SqMO) to produce 2,3-oxidosqualene (Tchen & Bloch, 1957; Willett *et al.*, 1967).

It has been proposed that the early epoxidation of squalene could have been achieved in the absence of O_2 (Raymond & Blankenship, 2004). The best candidate for the 'anaerobic' formation of 2,3-oxidosqualene is the utilization of HOOH or ROOH in a direct epoxidation reaction (Eq. 1). This reaction is simpler than the enzymatic version carried out by SQMO; indeed, the purpose of the enzymatic pathway is to convert O_2 into a bound peroxy group, and thus it requires reducing power supplied by NADPH (Eq. 2).

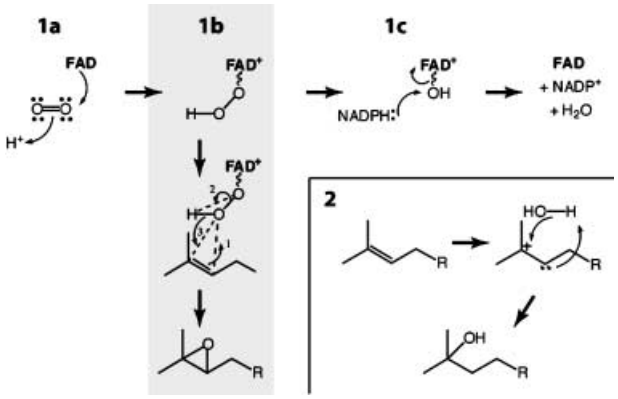

Fig. 6 Epoxidation of squalene by squalene monooxygenase (SqMO) (steps 1a–1c). (1a) O_2 forms an activated organic peroxy compound on flavin adenine dinucleotide (FAD). (1b) The terminal oxygen of the hydroperoxyflavin is inserted across the 2(3) double bond of squalene creating 2,3-oxidosqualene. (1c) FAD is regenerated; the FAD-bound oxygen is reduced to water with electrons from NADPH-cytochrome P450 reductase. Step (2) shows the abiotic or enzymatic hydration of the isoprene double bond by water. The –OH group is a nucleophile and attacks the 3° position.

$$H_2O_2 + \text{alkene} \Rightarrow \text{epoxide} + H_2O \quad (1)$$

$$O_2 + \text{NADPH} + H^+ + \text{alkene} \Rightarrow \text{epoxide} + \text{NADP}^+ + H_2O \quad (2)$$

In synthetic organic chemistry, peroxy acids commonly are used to epoxidize alkenes. Inserting oxygen across a carbon–carbon double bond requires that two C become oxidized, while the O must be above the O^{2-} oxidation state so it can be reduced. More importantly, peroxides contain a critical OR^- leaving group: the terminal O of HOOR acts both as an electrophile (with loss of OR^-) and as a nucleophile (with loss of H^+) during epoxidation; as such, H_2O or other O^{2-} hydroxyl donors that lack favourable leaving groups are too reduced to be adequate substitutes.

Some researchers have suggested that H_2O_2 may have been an important oxidant before the rise of molecular oxygen (Kasting, 1984; Kasting *et al.*, 1985). Borda *et al.* (2001) demonstrated H_2O_2 production up to concentrations of ~100 μM from a reaction of pyrite and water under anoxic conditions both in the dark and under visible light. Therefore it is at least possible that H_2O_2 or organic peroxides could have been used as an 'anaerobic' means to oxidize squalene in early versions of this reaction (Fig. 6, step 1b).

Could this step be achieved using H_2O? An alternative proposal for anaerobic activation of squalene invokes hydration of the terminal double bond of squalene (Raymond & Blankenship, 2004). However, this reaction would not place the resulting –OH group on the correct C_3 position of squalene (Fig. 6, inset 2), and thus it is not a viable mechanism to catalyse the formation of cyclic end-products. During ring opening of squalene epoxide (section *2.2.2*), the cation is transferred to the tertiary C_2 carbon, leaving the –OH attached to C_3. In contrast, direct hydration (abiotically or by enzymatic hydroxylation) of the double bond would place the -OH at C_2 by nucleophilic attack. This quenches the possibility of initiating a cyclization reaction, and the reaction would come to a dead end. Present evidence is therefore consistent with the interpretation that a 2,3-oxidosqualene substrate reflects oxidizing conditions, whether by O_2 or by ROOH.

3.3.4 Fitness of polycyclic triterpenoids

This paper focuses on the evolution of polycyclic tritrepenoids from a biosynthetic perspective. Such reductionist biology does not directly assess the linkages between specific molecular structures and Darwinian fitness (Kreitman & Akashi, 1995; Benner & Ricardo, 2005). Predicting the fitness of these molecules throughout their evolutionary history is difficult, because we are predisposed to certain assumptions about how such molecules ought to function. These assumptions often are based on studies of a limited number of model organisms.

Generally it is thought that the primary function of steroids and hopanoids is the stabilization and condensation of fatty acid membranes (Rohmer *et al.*, 1979; Dahl *et al.*, 1980; Kannenberg *et al.*, 1980; Bloch, 1983; Rohmer *et al.*, 1984; Ourisson *et al.*, 1987; Kannenberg & Poralla, 1999). However, there are at least several cases where this interpretation appears to be oversimplified. Examples include the hopanoid-dominated vesicles found in Frankia (Berry *et al.*, 1993) and liquid-ordered sterols in lipid rafts (Simons & Ikonen, 1997; Bacia *et al.*, 2005). Nevertheless, there may be general properties of certain classes of polycyclic triterpenoids that tend to make them more suitable or less suitable for useful biological functions. In particular, the dammaranoids may not have been very useful molecules if the primary biosynthetic product was dammaradiene and lacked amphiphilic properties. The poor fitness of dammaradiene may have prompted rapid evolution of the cyclase, D, towards other products (steroids and hopanoids), followed by the extinction of dammaranoid cyclases. If this was the case, it may also explain why dammaranes currently do not appear in the geological record until the Mesozoic, where they are associated with the evolution of angiosperms. It is presumed that this is an entirely independent (convergent) evolution of a dammaranoid cyclase in plants. Plants are master chemists that produce wide arrays of antagonistic products, including hundreds of diverse terpenoids, to ward off herbivores and parasites. Dammaranoid cyclases therefore may have evolved in plants as the biosynthetic products again became useful.

4. CONCLUSIONS

Here we proposed evolutionary models for the history of polycyclic triterpenoids using the approaches of amino acid sequence analysis, minimum evolution, and biochemical energetics. Although speculative, these models help constrain inferences that can be made from ancient geological biomarker data. The models also provide a number of testable hypotheses. Two major questions guided this work. Can we construct a reasonable evolutionary history for these biosynthetic reactions by dividing the chemistry of the cyclization reaction into a set of independent characters? Does such a history constrain the utilization of molecular oxygen by the organisms that synthesize steroids? The evolutionary models are based on four principal observations:

1 Three characters can be defined to describe the evolution of the major classes of polycyclic triterpenoids: (i) the number of anti-Markovnikov ring closures, (ii) CCC or CBC conformation of the preformed substrate, and (iii) the use of squalene or 2,3-oxidosqualene as the substrate for cyclization. These characters are each controlled by distinct parts of the enzymes and represent independent degrees of freedom during natural selection.

2 Four competing phylogenies constructed from these characters could reasonably describe the evolution of polycyclic triterpenoids. Conservation of critical amino acid residues suggests that malabaricanoid cyclases could have been the most ancient and would have been succeeded by dammaranoid cyclases. The two most parsimonious trees for the ensuing radiation of all other classes of cyclases – yielding steroids, hopanoids, isoarborinol, and tetrahymanol – both suggest that hopanoid cyclases and steroid cyclases diverged independently from a common ancestor and may not be direct descendents of each other.

3 The transition from CCC to CBC prefolded substrates marks a major geologically detectable divergence in the history of polycyclic triterpenoids. However, this change does not necessarily imply the simultaneous adoption of an O_2-requiring biosynthetic pathway. The amino acid residues responsible for folding the substrate are not involved in the catalytic protonation of 2,3-oxidosqualene, and therefore the order of evolution of these characters remains unresolved. Insight into this problem may come from the discovery of new cyclases that preserve either the xxDC motif with a CCC-fold or alternatively a DxDD motif with a CBC-fold.

4 All known producers of sterols create the substrate 2,3-oxidosqualene by employing SqMO, NADPH, and molecular oxygen. Reasonable anaerobic alternatives to this process include the direct epoxidation of squalene using hydrogen peroxide or other organic peroxy acids, neither of which would require any additional reducing power from NADPH. Mechanisms that invoke hydration are not viable because the –OH group would be added to the incorrect position on squalene, prohibiting cyclization.

Future work will be able to test the hypotheses proposed here in several different ways. Improvements to lipid biomarker data obtained from well-preserved sedimentary rocks would allow a direct assessment of the historical record of polycyclic triterpenoids. We acknowledge, however, that the biomarker record (which requires thermally immature rocks) may not extend deep enough into Earth history to capture the early evolution of the polycyclic triterpenoid biosynthetic pathway. Other profitable approaches include the discovery of additional triterpenoid cyclase enzymes. For example, many modern anoxic environments contain malabaricanoids (e.g. Behrens *et al.*, 1999; Werne *et al.*, 2000). Therefore a malabaricanoid cyclase must still exist and is discoverable. We can predict that this enzyme should not have the anti-Markovnikov carbocation-propagating residues homologous to F601 and F605 (SHC numbering). Approaches that integrate genomics, experimental genetics, biochemistry, and modern-day environmental samples may provide a useful synthesis of information as a means to test the evolutionary hypotheses presented here.

ACKNOWLEDGEMENTS

We thank Jochen Brocks, Andrew Knoll, Tanja Bosak, Joe Kirschvink, John Grotzinger, and Meytal Budin Higgins for helpful discussions. We thank Dianne Newman and Kurt Konhauser for editorial handling, and we gratefully acknowledge the comments of Roger Summons and an anonymous reviewer. This work was supported by the David and Lucille Packard Foundation (A.P), the Agouron Institute (W.W.F), and by Harvard University. Funding for genomic sequencing and the availability of these data in the public domain are supported by The Institute for Genome Research (TIGR) and the Department of Energy (DOE).

REFERENCES

Abe I, Rohmer M, Prestwich GD (1993) Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chemical Reviews* **93**, 2189–2206.

Alexander K, Akhtar M, Boar RB, McGhie JF, Barton DHR (1972) The removal of the 32-carbon atom as formic acid in cholesterol biosynthesis. *Journal of the Chemical Society, Chemical Communications* 383–385.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Anding C, Rohmer M, Ourisson G (1976) Nonspecific biosynthesis of hopane triterpenes in a cell-free system from *Acetobacter rancens*. *Journal of the American Chemical Society* **98**, 1274–1275.

Bacia K, Schwille P, Kurzchalia T (2005) Sterol structure determines the separation of phases and the curvature of liquid-ordered phase in model membranes. *Proceedings of the National Academy of Sciences (USA)* **102**, 3272–3277.

Barton DHR, Jarman T, Watson KC, Widdowson DA (1975) Investigations on the biosynthesis of steroids and terpenoids. Part XII. Biosynthesis of 3β-hydroxy-triterpenoids and – steroids from (3*S*)-2,3-epoxy-2,3-dihydrosqualene. *Journal of the Chemical Society Perkin Transactions* **I**, 1134–1138.

Behrens A, Schaeffer P, Bernasconi S, Albrecht P (1999) 17(*E*)-13α(*H*)-Malabarica-14(27), 17, 21-triene, an unexpected tricyclic hydrocarbon in sediments. *Organic Geochemistry* **30**, 379–383.

Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ (2004) Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120.

Benner SA (2003) Interpretive proteomics – finding biological meaning in genome and proteome databases. *Advances in Enzyme Regulation* **43**, 271–359.

Benner SA, Ricardo A (2005) Planetary systems biology. *Molecular Cell* **17**, 471–472.

Berry AM, Harriott OT, Moreau RA, Osman SF, Benson DR, Jones AD (1993) Hopanoid lipids compose the *Frankia* vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proceedings of the National Academy of Sciences (USA)* **90**, 6091–6094.

Bird CW, Lynch JM, Pirt SJ, Reid WW (1971a) Steroids and squalene in *Methylococcus capsulatus* grown on methane. *Nature* **230**, 473–474.

Bird CW, Lynch JM, Pirt SJ, Reid WW (1971b) Identification of hop-22(29)-ene in prokaryotic organisms. *Tetrahedron Letters* **34**, 3189–3190.

Bloch K (1983) Sterol structure and membrane function. *CRC Critical Reviews in Biochemistry* **14**, 47–92.

Bode HB, Zeggel B, Silakowski B, Wenzel SC, Hans R, Müller R (2003) Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Molecular Microbiology* **47**, 471–481.

Borda MJ, Elsetinow AR, Schoonen MA, Strongin DR (2001) Pyrite-induced hydrogen peroxide formation as a driving force in the evolution of photosynthetic organisms on an early Earth. *Astrobiology* **1**, 283–288.

Bravo J, Perzl M, Hartner T, Kannenberg EL, Rohmer M (2001) Novel methylated triterpenoids of the gammacerane series from the nitrogen-fixing bacterium *Bradyrhizobium japonicum* USDA 110. *European Journal of Biochemistry FEBS* **268**, 1323–1331.

Brocks JJ, Logan GA, Buick R, Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036.

Brocks JJ, Buick R, Logan GA, Summons RE (2003a) Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia. *Geochemica et Cosmochemica Acta* **67**, 4289–4319.

Brocks JJ, Buick R, Summons RE, Logan GA (2003b) A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochemica et Cosmochemica Acta* **67**, 4321–4335.

Brocks JJ, Love GD, Snape CE, Logan GA, Summons RE, Buick R (2003c) Release of bound aromatic hydrocarbons from late Archean and Mesoproterozoic kerogens via hydropyrolysis. *Geochimica et Cosmochimica Acta* **67**, 1521–1530.

Dahl JS, Dahl CE, Bloch K (1980) Sterols in membranes: growth characteristics and membrane properties of *Mycoplasma capricolum* culutured on cholesterol and lanosterol. *Biochemistry* **19**, 1467–1472.

Dastillung M, Albrecht P, Ourisson G (1980) Aliphatic and polycyclic alcohols in sediments: hydroxylated derivatives of hopane and 3-methylhopane. *Journal of Chemical Research Synopses* **168–169**, M2353–2374.

Douka E, Koukkou A, Drainas C, Grosdemange-Billard C, Rohmer M (2001) Structural diversity of the triterpenic hydrocarbons from the bacterium *Zymomonas mobilis*: the signature of defective squalene cyclization by the squalene/hopene cyclase. *FEMS Microbiology Letters* **199**, 247–251.

Faquhar J, Bao H, Thiemens MH (2000) Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–759.
Feil C, Sussmuth R, Jung G, Poralla K (1996) Site-directed mutagenesis of putative active-site residues in squalene-hopene cyclase. *European Journal of Biochemistry* **242**, 51–55.
Fischer WW, Summons RE, Pearson A (2005) Targeted genomic detection of biosynthetic pathways: anaerobic production of hopanoid biomarkers by a common sedimentary microbe. *Geobiology* **3**, 33–40.
Flesch G, Rohmer M (1988) Prokaryotic triterpenoids. The biosynthesis of the bacteriohopane skeleton: formation of isoprenic units from two distinct acetate pools and a novel type of carbon/carbon linkage between a triterpene and D-ribose. *European Journal of Biochemistry* **175**, 405–411.
Frimmel HE (2005) Archaean atmospheric evolution: evidence from the Witwatersrand gold fields, South Africa. *Earth Science Reviews* **70**, 1–46.
Füll C, Poralla K (2000) Conserved Tyr residues determine functions of *Alicyclobacillus acidocaldarius* squalene-hopene cyclase. *FEMS Microbiology Letters* **183**, 221–224.
Härtner T, Straub KL, Kannenberg E (2005) Occurence of hopanoid lipids in anaerobic *Geobacter* species. *FEMS Microbiology Letters* **243**, 59–64.
Hauke V, Graff R, Wehrung P, Trendel JM, Albrecht P (1992a) Novel triterpene-derived hydrocarbons of arborane/fernane series in sediments: Part I. *Tetrahedron* **48**, 3915–3924.
Hauke V, Graff R, Wehrung P, Trendel JM, Albrecht P, Riva A, Hopfgartner G, Gulacar FO, Buchs A, Eakin PA (1992b) Novel triterpene-derived hydrocarbons of the arborane/fernane series in sediments: Part II. *Geochimica et Cosmochimica Acta* **56**, 3595–3602.
Hauke V, Adam P, Trendel JM, Albrecht P, Schwark L, Vliex M, Hagemann H, Puttmann W (1995) Isoarborinol through geological times: evidence for its presence in the Permian and Triassic. *Organic Geochemistry* **23**, 91–93.
Hoshino T, Sato T (2002) Squalene-hopene cyclase: catalytic mechanism and substrate recognition. *Chemical Communications* **4**, 291–301.
Hoshino T, Kouda M, Abe T, Sato T (2000) Functional analysis of Phe605, a conserved aromatic amino acid in squalene-hopene cyclases. *Chemical Communications* **16**, 1485–1486.
Jaffe R, Hausmann KB (1995) Origin and early diagenesis of arborinone/isoarborinol in sediments of a highly productive freshwater lake. *Organic Geochemistry* **22**, 231–235.
Jahnke LL (1986) The effects of low oxygen on the synthesis of unsaturated fatty acids and sterols: implications for the evolution of eukaryotes. *Origins of Life and Evolution of the Biosphere* **16**, 317–318.
Jahnke LL, Klein HP (1983) Oxygen requirements for formation and activity of the squalene epoxidase in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **155**, 488–492.
Jahnke LL, Nichols PD (1986) Methyl sterol and cyclopropane fatty acid composition of *Methylococcus capsulatus* grown at low oxygen tensions. *Journal of Bacteriology* **167**, 238–242.
Jenson C, Jorgensen WL (1997) Computational investigations of carbenium ion reactions relevant to sterol biosynthesis. *Journal of the American Chemical Society* **119**, 10846–10854.
Kannenberg EL, Poralla K (1999) Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften* **86**, 168–176.
Kannenberg EL, Poralla K, Blume A (1980) A hopanoid from the thermoacidiphilic *Bacillus acidocaldarius* condenses membranes. *Naturwissenschaften* **67**, 458–459.
Kasting JF (1984) The evolution of the prebiotic atmosphere. *Origins of Life* **14**, 75–82.
Kasting JF, Holland HD, Pinto JP (1985) Oxidant abundances in rainwater and the evolution of atmospheric oxygen. *Journal of Geophysical Research* **90**, 10497–10510.
Kleemann G, Poralla K, Englert G, Kjosen H, Liaaen-Jensen S, Neunlist S, Rohmer M (1990) Tetrahymanol from the phototrophic bacterium *Rhodopseudomonas palustris*: first report of a gammacerane triterpene from a prokaryote. *Journal of General Microbiology* **136**, 2551–2553.
Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ (2005) The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences (USA)* **102**, 11131–11136.
Kreitman M, Akashi H (1995) Molecular evidence for natural selection. *Annual Review of Ecology and Systematics* **26**, 403–422.
Lenhart A, Weihofen WA, Pleschke AEW, Schultz (2002) Crystal structure of a squalene cyclase in complex with potential anticholesteremic drug Ro48-8071. *Chemistry and Biology* **9**, 639–645.
Merkofer T, Pale-Gosdemange C, Wendt KU, Rohmer M, Poralla K (1999) Altered product pattern of a squalene-hopene cyclase by mutagenesis of active site residues. *Tetrahedron Letters* **40**, 2121–2124.
Miller WL, Kalafer ME, Gaylor JL, Delwiche CV (1967) Investigation of the component reactions of oxidative sterol demethylation. Study of the aerobic and anaerobic processes. *Biochemistry* **6**, 2673–2678.
Neunlist S, Holst O, Rohmer M (1985) Prokaryotic triterpenoids. The hopanoids of the purple non-sulfur bacterium *Rhodomicrobium vannielii*: an aminotriol and its aminoacyl derivatives, N-tryptophanyl and N-ornithyl aminotriol. *European Journal of Biochemistry* **147**, 561–568.
Nytoft HP, Larsen B (2001) Triterpenoids and other organic compounds as markers of depositional conditions in the Baltic Sea deep basins during the Holocene. *Baltica* **14**, 95–107.
Ono T, Bloch K (1975) Solubilization and partial characterization of rat liver squalene epoxidase. *Journal of Biological Chemistry* **250**, 1571–1579.
Ourisson G, Nakatani Y (1994) The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chemistry and Biology* **1**, 11–23.
Ourisson G, Albrecht P, Rohmer M (1982) Predictive microbial biochemistry – from molecular fossils to prokaryotic membranes. *Trends in Biochemical Sciences* 236–239.
Ourisson G, Rohmer M, Poralla K (1987) Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annual Reviews of Microbiology* **41**, 301–333.
Page RDM, Holmes EC (1998) *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science, Massachusetts.
Pale-Gosdemange C, Feil C, Rohmer M, Poralla K (1998) Occurrence of cationic intermediates and deficient control during the enzymatic cyclization of squalene to hopanoids. *Angewandte Chemie International Edition (English)* **37**, 2237–2240.
Paton WF, Paul IC, Bajaj AG, Dev S (1979) Structure of malabaricaol. *Tetrahedron Letters* **43**, 4153–4154.
Pavlov AA, Kasting JF (2002) Mass-independent fractionation of sulfur isotopes in archean sediments: strong evidence for an anoxic archean atmosphere. *Astrobiology* **2**, 27–41.
Pearson A, Brocks JJ, Budin M (2003) Phylogenetic and biochemical evidence for sterol biosynthesis in the bacterium, *Gemmata obscuriglobus*. *Proceedings of the National Academy of Sciences (USA)* **100**, 15352–15357.
Perzl M, Muller P, Poralla K, Kannenberg EL (1997) Squalene-hopene cyclase from *Bradyrhizobium japonicum*: cloning, expression, sequence analysis and comparison to other triterpenoid cyclases. *Microbiology* **143**, 1235–1242.

Perzl M, Reipen IG, Schmitz S, Poralla K, Sahm H, Sprenger GA, Kannenberg EL (1998) Cloning of conserved genes from *Zymomonas mobilis* and *Bradyrhizobium japonicum* that function in the biosynthesis of hopanoid lipids. *Biochimica et Biophysica Acta* **1393**, 108–118.

Peters KE, Moldowan JM (1993) *The Biomaker Guide*. Prentice Hall, New York.

Pratt LM, Summons RE, Hieshima GB (1991) Sterane and triterpane biomarkers in the Precambrian Nonesuch Formation, North American Midcontinent Rift. *Geochimica et Cosmochimica Acta* **55**, 911–916.

Rajamani R, Gao J (2003) Balancing kinetic and thermodynamic control: the mechanism of carbocation cyclization by squalene cyclase. *Journal of the American Chemical Society* **125**, 12768–12781.

Rasmussen B, Buick R (1999) Redox state of the Archean atmosphere: evidence from detrital heavy minerals in ca. 3250-2750 Ma sandstones from the Pilbara Craton, Australia. *Geology* **27**, 115–118.

Raymond J, Blankenship RE (2004) Biosynthetic pathways, gene replacement and the antiquity of life. *Geobiology* **2**, 199–203.

Rohmer M, Bouvier P, Ourisson G (1979) Molecular evolution of biomembranes: structural equivalents and phylogenetic precursors of sterols. *Proceedings of the National Academy of Sciences (USA)* **76**, 847–851.

Rohmer M, Bouvier P, Ourisson G (1980) Non-specific lanosterol and hopanoid biosynthesis by a cell-free system from the bacterium *Methylococcus capsulatus*. *European Journal of Biochemistry* **112**, 557–560.

Rohmer M, Bouvier-Nave P, Ourisson G (1984) Distribution of hopanoid triterpenes in Prokaryotes. *Journal of General Microbiology* **130**, 1137–1150.

Rye R, Holland HD (1998) Paleosols and the evolution of atmospheric oxygen: a critical review. *American Journal of Science* **298**, 621–672.

Sato T, Hoshino T (1999) 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. *Bioscience Biotechnology Biochemistry* **63**, 2189–2198.

Sato T, Hoshino T (2001) Catalytic function of the residues of phenylalanine and tyrosine conserved in squalene-hopene cyclases. *Bioscience Biotechnology Biochemistry* **65**, 2233–2242.

Schouten S, Hoefs MJL, Damste JSS (2000) A molecular and stable carbon isotopic study of lipids in late Quaternary sediments from the Arabian Sea. *Organic Geochemistry* **31**, 509–521.

Schultz-Gasch T, Stahl M (2003) Mechanistic insights into oxidosqualene cyclizations through homology modeling. *Journal of Computational Chemistry* **24**, 741–753.

Shibuya M, Zhang H, Endo A, Shishikura K, Kushiro T, Ebizuka Y (1999) Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclases. *European Journal of Biochemistry* **266**, 302–307.

Simons K, Ikonen E (1997) Functional rafts in cell membranes. *Nature* **387**, 569–572.

Sinninghe Damsté JS, Rijpstra WIC, Schouten S, Fuerst JA, Jetten MSM, Strous M (2004) The occurrence of hopanoids in planctomycetes: implications for the sedimentary biomarker record. *Organic Geochemistry* **35**, 561–566.

Summons RE, Powell TG, Boreham CJ (1988a) Petroleum geology and geochimstry of the Middle Proterozoic McArthur Basin, Northern Australia: III. Composition of extractable hydrocarbons. *Geochimica et Cosmochimica Acta* **52**, 1747–1763.

Summons RE, Brassell SC, Eglinton G, Evans E, Horodyski RJ, Robinson N, Ward DM (1988b) Distinctive hydrocarbon biomarkers from fossiliferous sediments of the Late Proterozoic Walcott Member, Chuar Group, Grand Canyon, Arizona. *Geochimica et Cosmochimica Acta* **52**, 2625–2637.

Summons RE, Bradley AS, Jahnke LL, Waldbauer JR (2006) Steroids, triterpenoids, and molecular oxygen. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **361**, 951–968.

van Tamelen EE, Willett JD, Schwarz M, Nadeau R (1966) Nonenzymic laboratory cyclization of squalene 2,3-oxide. *Journal of the American Chemical Society* **88**, 5937–5938.

Tchen TT, Bloch K (1957) On the conversion of squalene to lanosterol in vitro. *Journal of Biochemistry* **226**, 921–930.

Thoma R, Schulz-Gasch T, D'Arcy B, Benz J, Aebi J, Dehmlow H, Hennig M, Stihle M, Ruf A (2004) Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. *Nature* **432**, 118–122.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Tippelt A, Jahnke L, Poralla K (1998) Squalene-hopene cyclase from *Methlyococcus capsulatus* (Bath): a bacterium producing hopanoids and steroids. *Biochimica et Biophysica Acta* **1391**, 223–232.

Wendt KU, Schultz GE (1998) Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure* **6**, 127–133.

Wendt KU, Poralla K, Schultz GE (1997) Structure and function of a squalene cyclase. *Science* **277**, 1811–1815.

Wendt KU, Schultz GE, Corey EJ (2000) Enzyme mechanisms for polycyclic triterpene formation. *Angewandte Chemie International Edition (English)* **39**, 2812–2833.

Werne JP, Hollander DJ, Behrens A, Shaeffer P, Albrecht P, Sinninghe Damsté JSS (2000) Timing of early diagenetic sulfurization of organic matter: a precursor-product relationship in Holocene sediments of the anoxic Cariaco Basin, Venezuela. *Geochimica et Cosmochimica Acta* **64**, 1741–1751.

Willett JD, Sharpless KB, Lord KE, van Tamelen EE, Clayton RB (1967) Squalene-2,3oxide, and intermediate in the enzymatic conversion of squalene to lanosterol and cholesterol. *Journal of Biological Chemistry* **242**, 4182–4191.

Woodward RB, Bloch K (1953) The cyclization of squalene in cholesterol synthesis. *Journal of the American Chemical Society* **75**, 2023–2024.

Xiong Q, Rocco F, Wilson W, Xu R, Ceruti M, Matsuda SPT (2005) Structure and reactivity of the dammarenyl cation: configurational transmission in triterpene synthesis. *Journal of Organic Chemistry* **70**, 5362–5375.

Xu R, Fazio GC, Matsuda SPT (2004) On the origins of triterpenoid skeletal diversity. *Phytochemistry* **65**, 261–291.

Yamamoto S, Bloch K (1970) Studies on squalene epoxidase of rat liver. *Journal of Biological Chemistry* **245**, 1670–1674.