# Biosynthesis of Triterpenoid Saponins in Plants

Kosmas Haralampidis, Miranda Trojanowska and Anne E. Osbourn

Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK,
e-mail: annie.osbourn@bbsrc.ac.uk

Many different plant species synthesise triterpenoid saponins as part of their normal programme of growth and development. Examples include plants that are exploited as sources of drugs, such as liquorice and ginseng, and also crop plants such as legumes and oats. Interest in these molecules stems from their medicinal properties, antimicrobial activity, and their likely role as determinants of plant disease resistance. Triterpenoid saponins are synthesised via the isoprenoid pathway by cyclization of 2,3-oxidosqualene to give primarily oleanane ($\beta$-amyrin) or dammarane triterpenoid skeletons. The triterpenoid backbone then undergoes various modifications (oxidation, substitution and glycosylation), mediated by cytochrome P450-dependent monooxygenases, glycosyltransferases and other enzymes. In general very little is known about the enzymes and biochemical pathways involved in saponin biosynthesis. The genetic machinery required for the elaboration of this important family of plant secondary metabolites is as yet largely uncharacterised, despite the considerable commercial interest in this important group of natural products. This is likely to be due in part to the complexity of the molecules and the lack of pathway intermediates for biochemical studies. Considerable advances have recently been made, however, in the area of 2,3-oxidosqualene cyclisation, and a number of genes encoding the enzymes that give rise to the diverse array of plant triterpenoid skeletons have been cloned. Progress has also been made in the characterisation of saponin glucosyltransferases. This review outlines these developments, with particular emphasis on triterpenoid saponins.

**Keywords.** Saponins, Triterpenoids, Sterols, 2,3-Oxidosqualene cyclases, Glycosyltransferases



Advances in Biochemical Engineering/
Biotechnology, Vol. 75
Managing Editor: Th. Scheper

## List of Abbreviations

$\alpha$AS $\alpha$-Amyrin synthase
$\beta$AS $\beta$-Amyrin synthase
CS Cycloartenol synthase
DS Dammarenediol synthase
EST Expressed sequence tag
LS Lanosterol synthase
LuP Lupeol synthase
OSC Oxidosqualene cyclase
PCR Polymerase chain reaction
SC Squalene-hopene cyclase

## 1 Introduction

Collectively plants synthesise a diverse array of secondary metabolites, either as part of normal growth and development or in response to pathogen attack or stress. By definition, secondary metabolites are regarded as "luxury items" that are not required for growth and reproduction of the plant, at least under defined conditions. Since the ability to synthesise particular classes of secondary metabolites is restricted to certain plant groups, these compounds are clearly not essential for survival. However they may be important in conferring selective advantages, for example by suppressing the growth of neighbouring plants or by protecting against pests, pathogens and stress [1–5]. They may also have subtle physiological roles in plants, which are as yet uncharacterised. In addition to their natural roles plant secondary metabolites also represent a vast resource of complex molecules that are valued and exploited by man for their pharmacological and other properties [1].

Saponins are an important group of plant secondary metabolites that are widespread throughout the Plant Kingdom [4, 6–9]. The name saponin is derived from *sapo*, the Latin word for soap, since these molecules have surfactant properties and give stable, soap-like foams in aqueous solution. Chemically, the term saponin has become accepted to define a group of structurally diverse molecules that consists of glycosylated steroids, steroidal alkaloids and triterpenoids (Fig. 1). These secondary metabolites often occur in plants as complex mixtures, and saponin content and composition may vary markedly depending on the genetic background of the plant material, the tissue type, the age and physiological state of the plant and environmental factors [6–10].

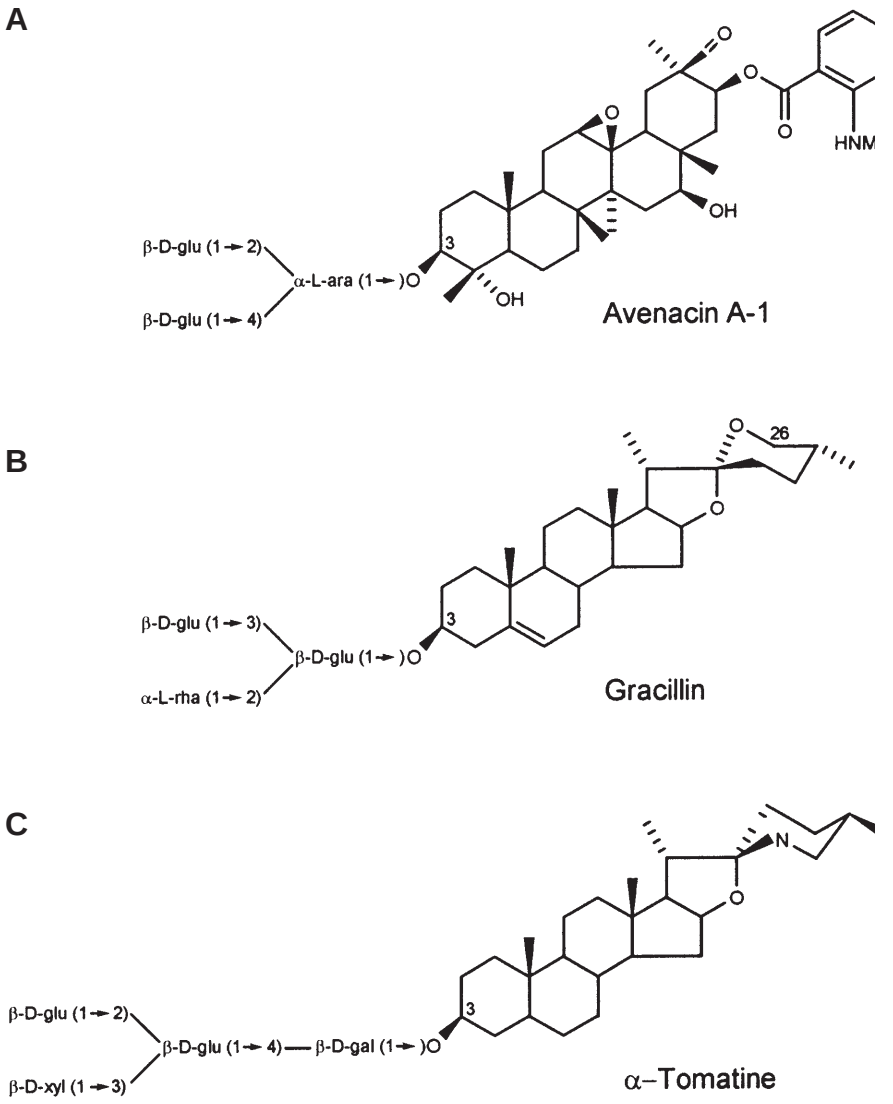


**Fig. 1 A–C.** Examples of different classes of saponins: **A** the triterpenoid saponin avenacin A-1 from roots of *Avena* spp; **B** the steroidal saponin gracillin, from *Costus speciosus*; **C** the steroidal glycoalkaloid $\alpha$-tomatine from tomato (*Lycopersicon* spp.)

Saponins have been variously attributed with a diverse range of properties, some of which include both beneficial and detrimental effects on human health, piscidical, insecticidal and molluscicidal activity, allelopathic action, antinutritional effects, sweetness and bitterness, and as phytoprotectants that defend plants against attack by microbes and herbivores [2–11]. A more detailed understanding of the biochemical pathways and enzymes involved in saponin

biosynthesis will facilitate the development of plants with altered saponin content. In some cases enhanced levels of saponins or the synthesis of novel saponins may be desirable (for example, for drug production [9, 12–15] or improved disease resistance [4, 6, 10, 16]) while for other plants reduction in the content of undesirable saponins would be beneficial (for example, for legume saponins that are associated with antifeedant properties in animal feed [8]). This review is concerned with recent progress that has been made in the characterisation of the enzymes and genes involved in the synthesis of these complex molecules, and focuses on triterpenoid saponins.

## 2 Cyclization of 2,3-Oxidosqualene – The First Committed Step in Triterpenoid Biosynthesis

The first committed step in the synthesis of triterpenoid saponins involves the cyclisation of 2,3-oxidosqualene to give one of a number of different potential products [9]. Most plant triterpenoid saponins are derived from oleanane or dammarane skeletons, although lupanes are also common [9]. This cyclisation event forms a branchpoint with the sterol biosynthetic pathway, in which 2,3-oxidosqualene is cyclised to lanosterol (in animals and fungi) or to cycloartenol (in plants) (Fig. 2). Sterols are important membrane constituents and also serve as precursors for hormone biosynthesis.

The cyclisation, rearrangement and deprotonation reactions leading to the different products shown in Fig. 2 (originally proposed by the "biogenetic isoprene rule") are well established [17–20]. Enzymatic cyclisation of 2,3-oxidosqualene into sterols proceeds in the "chair-boat-chair" conformation to yield the C-20 protosteryl cation, which is then converted to cycloartenol or lanosterol. These cyclisation events are catalysed by the 2,3-oxidosqualene cyclases (OSCs) cycloartenol synthase (CS) and lanosterol synthase (LS), respectively. Triterpenoid synthesis, on the other hand, involves cyclisation of the "chair-chair-chair" conformation of the substrate to give the tetracyclic dammarenyl cation. This cation may then be converted to dammarene-like triterpenoids by the OSC dammarenediol synthase (DS), or may undergo further rearrangements leading to the formation of pentacyclic triterpenoids derived from lupeol, $\beta$-amyrin and $\alpha$-amyrin (Fig. 2). The 2,3-oxidosqualene cyclases (OSCs) that mediate these different cyclisation events are listed in Table 1.

**Table 1.** 2,3-Oxidosqualene cyclases

| Enzyme | Abbreviation | EC Number | Organism |
|---|---|---|---|
| Lanosterol synthase | LS | EC 5.4.99.7 | Animals, Fungi |
| Cycloartenol synthase | CS | EC 5.4.99.8 | Plants |
| Dammarenediol synthase | DS | – | Plants |
| $\alpha$-Amyrin synthase | $\alpha$AS | EC 5.4.99 | Plants |
| $\beta$-Amyrin synthase | $\beta$AS | EC 5.4.99 | Plants |
| Lupeol synthase | LuS | EC 5.4.99 | Plants |

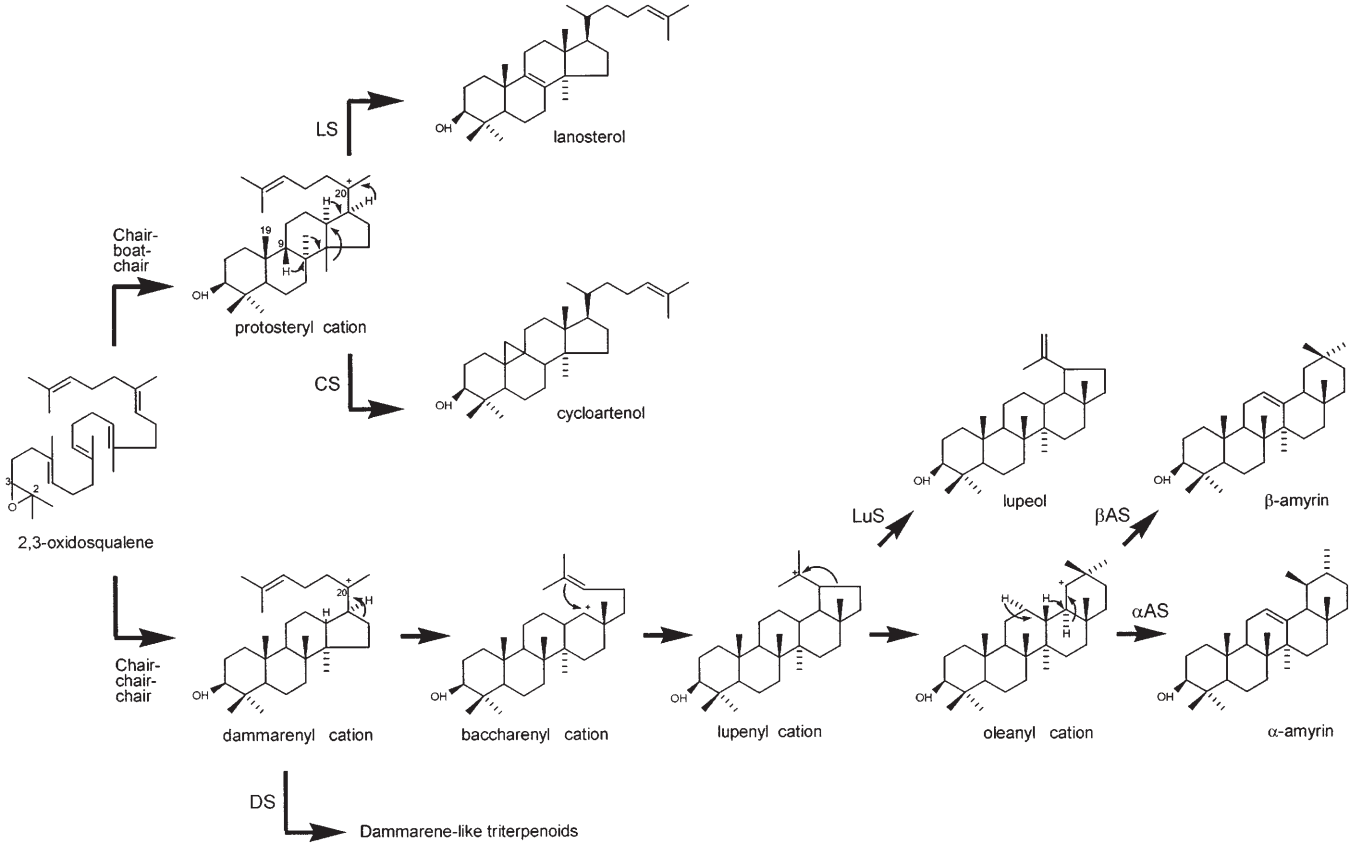


**Fig. 2.** Cyclization of 2,3-oxidosqualene to sterols and triterpenoids. The 2,3-oxidosqualene cyclase enzymes that catalyse the formation of the different products are indicated: LS, lanosterol synthase; CS, cycloartenol synthase; DS, dammarenediol synthase; LuS, lupeol synthase; $\beta$AS, $\beta$-amyrin synthase; $\alpha$AS, $\alpha$-amyrin synthase

The mechanisms by which 2,3-oxidosqualene is cyclised to a diverse range of products has been a source of intrigue for nearly half a century, and the enzymes that catalyse these reactions are of great interest both to biochemists and to industry. LS enzymes have commercial importance as targets for the development of antifungal [21] and cholesteremic drugs [21, 22]. Since the cyclisation of 2,3-oxidosqualene to sterols and triterpenoids represents a branchpoint between primary and secondary metabolism, plant OSCs are also attractive tools for investigating the regulation of synthesis and the physiological role of triterpenoids, and potentially for manipulation of sterol and triterpenoid content [23–26]. Recent progress in the purification of OSCs and in the cloning and analysis of the corresponding genes has given us substantial insight into the relationship between the nature of the cyclisation event and enzyme structure, and distinct subgroups of enzymes that mediate the conversion of 2,3-oxidosqualene to different cyclisation products are now emerging. Advances in this area are summarised below.

### 2.1
### Resolution of Cyclase Activities Required for Sterol and Triterpenoid Biosynthesis

A key question in understanding triterpenoid biosynthesis has centred around whether the generation of different cyclisation products from 2,3-oxidosqualene involves distinct oxidosqualene cyclase enzymes, or whether these reactions may be mediated by a single enzyme, the product specificity of which may be determined by protein modification or by factors such as electrolyte concentration [23, 27, 28]. In pea (*Pisum sativum*), $\beta$-amyrin production is very active during development and just after germination, while sterol biosynthesis increases several days after germination [23]. Similar changes in triterpenoid and steroid biosynthesis in developing seed have been reported for the monocot *Sorghum bicolor* suggesting that this may be a common phenomenon in different plant species [29], although the significance of this dramatic switch between sterol and triterpenoid synthesis is unclear. The levels of $\beta$-amyrin synthase ($\beta$AS) and cycloartenol synthase (CS) activities in germinating pea seedlings alter during development in parallel with the changes in sterol and triterpenoid content, suggesting that the two enzymes are likely to be distinct proteins [27].

CS and $\beta$AS enzymes have been fully purified from seedlings of pea [27, 30, 31] and also from microsomes of cell suspension cultures of another plant species, *Rabdosia japonica* [32]. The purified pea CS and $\beta$AS enzymes were identified as 55-kDa and 35-kDa proteins, respectively [27, 30, 31]. Similarly CS and $\beta$AS enzymes purified from *R. japonica* were also found to be distinct protein species with molecular masses of 54 kDa and 28 kDa, respectively [32]. The two classes of OSC also show differences in sensitivity to inhibitors and detergents [27, 32]. Taken together, this information suggested that the two cyclisation reactions were carried out by different enzymes. However, in the absence of amino acid sequence information the possibility remained that the enzymes represented different forms of a single gene product. The cloning and characterisation of genes encoding OSCs was required to resolve this.

## 2.2
## Cloning of 2,3-Oxidosqualene Cyclases

### 2.2.1
### *OSCs Required for Sterol Biosynthesis*

In general, purification of OSCs in sufficient quantity for amino acid sequence determination or antibody production has proved to be difficult due to the small amounts of protein and the problems of obtaining the solubilised enzyme in an active state [20, 27, 33, 34], and so attempts to isolate cDNA clones or genes encoding OSCs by these routes have met with only limited success. However, amino acid sequence information from purified OSCs has been used successfully in reverse genetics approaches to isolate cloned cDNAs for LS from animals and fungi. For example, the cDNA encoding rat LS was cloned in this way [34, 35]. The LS genes from *Candida albicans* and *Saccharomyces cerevisiae* have both been cloned using a different strategy that involved genetic complementation of an LS-deficient (*erg7*) mutant of *S. cerevisiae* [36–38].

Yeast does not synthesise cycloartenol or triterpenes, and so approaches to clone plant OSCs by complementation in *S. cerevisiae* are not feasible because of the lack of appropriate mutants. However, LS-deficient yeast mutants accumulate high levels of 2,3-oxidosqualene, favouring the synthesis of novel cyclisation products generated by heterologous expression of OSCs. The absence of lanosterol also facilitates analysis of the reaction products. Corey and co-workers isolated a cDNA encoding *Arabidopsis thaliana* CS by transforming a plant cDNA expression library into such a yeast mutant and screening protein preparations derived from pools of transformants for the ability to synthesise cycloartenol by TLC [39].

The availability of an increasing amount of DNA sequence information for LS and CS enzymes enabled regions of highly conserved amino acids to be identified in their predicted amino acid sequences. The cDNAs encoding LS from yeast, humans and rat have been cloned by polymerase chain reaction (PCR) amplification using degenerate oligonucleotide primers corresponding to such regions [40, 41]. The *P. sativum* CS cDNA was also isolated in this way [42]. The functions of cDNAs that are predicted to encode OSCs have generally been confirmed by expression in LS-deficient yeast strains. Cycloartenol does not normally accumulate in sufficient amounts to be detected by TLC or HPLC in yeast extracts (probably due to metabolic conversion of cycloartenol by the yeast strain [12, 43]) and so assays for cycloartenol synthase activity are usually carried out in vitro using [$^3$H]oxidosqualene as the substrate.

### 2.2.2
### *Triterpenoid Cyclases*

An important step was made when the conservation between the predicted amino acid sequences of cloned and characterised LS and CS enzymes was used to isolate two cDNA clones that were predicted to encode OSCs from *Panax ginseng* [12]. Expression of the full-length cDNAs in yeast indicated that one of the

clones encoded CS, while the other encoded βAS. Triterpenoid products generated by heterologous expression of OSCs in yeast are not readily metabolised, and so can be detected by analysis of cell extracts by TLC or reverse-phase HPLC. The deduced amino acid sequence of the *P. ginseng* βAS is 60% identical to that of the CS enzyme, suggesting that the enzymes may share a common ancestral origin. A second βAS has also been cloned from *P. ginseng* [14]. The subsequent cloning of cDNAs encoding α-/β-amyrin cyclase enzymes from pea [44], LuS enzymes from *Olea europaea* and *Taraxacum officinale* [45], and multifunctional triterpene synthases from *Arabidopsis thaliana* [46–48], confirmed that there is overall structural relatedness between the LS and CS enzymes of sterol biosynthesis and the OSCs that mediate the cyclisation of 2,3-oxidosqualene to triterpenoids. However, the triterpenoid cyclases are clearly distinct from LS and CS enzymes, and form discrete subgroups within the OSC superfamily.

### 2.2.3
### *OSC Gene Families*

Benveniste et al. have carried out the first comprehensive analysis of OSCs in a single plant species by screening *A. thaliana* EST databases and cDNA libraries and have identified 3 cDNAs encoding OSCs [47]. The function of these has been investigated by expression in yeast. One (ATLUP1) was identical to an LuS isolated from *A. thaliana* by Matsuda and co-workers [46], which synthesises primarily lupeol in yeast [46, 47]. ATLUP1 also forms other minor triterpene products including β-amyrin in yeast [46, 47, 49] and because it is more closely related to the *P. ginseng* βAS enzymes than to the LuS enzymes from *Olea europaea* and *Taraxacum officinalis* it has been suggested that in the plant the primary product of ATLUP1 may be β- or α-amyrin or other triterpenoids [47, 49]. The second cDNA (ATLUP2) encoded a multifunctional enzyme that catalysed the production of lupeol, β- and α-amyrin in the ratio 15:55:30 [47, 48]. A third predicted OSC (ATPEN1) was also introduced into yeast but no triterpenoid products were detected. A survey of *Arabidopsis* genomic sequence information identified five genes belonging to the subfamily of OSCs that contained ATLUP1 and ATLUP2, and seven genes that were closely related to ATPEN1. The functions of these triterpenoid cyclases in *A. thaliana* are as yet unknown. *A. thaliana* does not appear to synthesise saponins, but α- and β-amyrin and lupeol are present in extracts of leaves and callus of the plant [47]. It is possible that triterpenoids and their derivatives may play important roles in plant growth and development.

## 2.3
## The Relationship Between Structure and Function

OSCs that have been functionally characterised by expression in yeast are listed in Table 2, and the relatedness between the amino acid sequences of these enzymes is illustrated in Fig. 3. The conserved features of this OSC superfamily and the differences between them that may confer product specificity are considered below.

**Table 2.** 2,3-Oxidosqualene cyclases that have been cloned and their function confirmed by expression in yeast

| Organism | Enzyme | Abbreviation | EMBL/Gene Bank AC | Ref. |
|---|---|---|---|---|
| *Rattus norvegicus* | Lanosterol synthase | OSC | U31352 | 34 |
| *Saccharomyces cerevisiae* | Lanosterol synthase | ERG7 | U04841 | 40 |
| *Candida albicans* | Lanosterol synthase | ERG7 | L04305 | 38 |
| *Arabidopsis thaliana* | Cycloartenol synthase | CAS1 | U02555 | 39 |
| *Pisum sativum* | Cycloartenol synthase | PSX | D89619 | 42 |
| *Panax ginseng* | Cycloartenol synthase | PNX | AB009029 | 12 |
| *Panax ginseng* | $\beta$-Amyrin synthase | PNY | AB009030 | 12 |
| *Panax ginseng* | $\beta$-Amyrin synthase | PNY2 | AB014057 | 14 |
| *Pisum sativum* | $\beta$-Amyrin synthase | PSY | AB034802 | 44 |
| *Pisum sativum* | Mixed amyrin synthase[a] | PSM | AB034803 | 44 |
| *Olea europaea* | Lupeol synthase | OEW | AB025343 | 45 |
| *Taraxacum officinale* | Lupeol synthase | TRW | AB025345 | 45 |
| *Arabidopsis thaliana* | Lupeol synthase[b] | ATLUP1 | U49919 | 46 |
| *Arabidopsis thaliana* | Multifunctional[c] | ATLUP2 | AF003472 | 47, 48 |

[a] Primary products $\alpha$- and $\beta$-amyrin (60%:40%) and also other minor triterpenes [44].
[b] Synthesises mainly lupeol but also produces at least five other minor triterpenoids [46, 47, 49].
[c] Primary products $\beta$-amyrin, $\alpha$-amyrin and lupeol (55%:30%:15%) [47, 48].

### 2.3.1 *Conserved Features*

Currently there is no experimentally determined three-dimensional structural information available for OSCs, although studies with a related enzyme, squalene-hopene cyclase (SC; EC 5.4.99.7) have proved informative. SCs are involved in the direct cyclisation of squalene to pentacyclic triterpenoids known as hopanoids, which play an integral role in membrane structure in prokaryotes [51]. A number of SC genes have been cloned from bacteria [52–54]. The SC and OSC enzymes have related predicted amino acid sequences, and so should have similar spatial structures [55]. The crystal structure of recombinant SC from the Gram-positive bacterium *Alicyclobacillus acidocaldarius* has established that the enzyme is dimeric [55]. Each subunit consists of two $\alpha$-$\alpha$ barrel domains that assemble to form a central hydrophobic cavity [55, 56].

The activity of 2,3-oxidosqualene cyclases is associated with microsomes, indicating their membrane-bound nature. However, the predicted amino acid sequences of these enzymes generally lack signal sequences and obvious transmembrane domains. Addition of hydrophobic membrane-localising regions to OSCs during evolution may have removed selection pressures that maintained alternate mechanisms for membrane localisation [33]. Consistent with this, there is a non-polar plateau on the surface of the *A. acidocaldarius* SC enzyme which is believed to be immersed in the centre of the membrane. The squalene substrate for SC is likely to diffuse from the membrane interior into the central cavity of the enzyme via this contact region [55, 56].

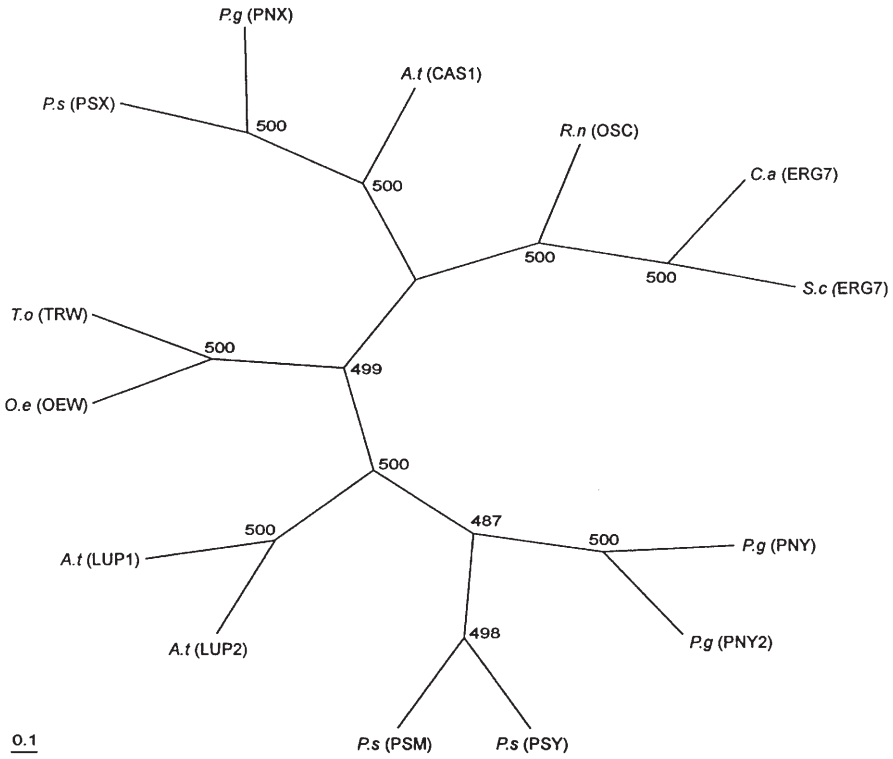


**Fig. 3.** Relatedness between deduced amino acid sequences of members of the OSC superfamily. *P.s* (PSX), *Pisum sativum* cycloartenol synthase (D89619); *P.g* (PNX), *Panax ginseng* cycloartenol synthase (AB009029); *A.t* (CAS1), *Arabidopsis thaliana* cycloartenol synthase (U02555); *R.n* (OSC), *Rattus norvegicus* lanosterol synthase (U31352); *C.a* (ERG7), *Candida albicans* lanosterol synthase (L04305); *S.c* (ERG7), *Saccharomyces cerevisiae* lanosterol synthase (U04841); *P.g* (PNY), *Panax ginseng* $\beta$-amyrin synthase (AB009030); *P.g* (PNY2), *Panax ginseng* $\beta$-amyrin synthase (AB014057); *P.s* (PSY), *Pisum sativum* $\beta$-amyrin synthase (AB034802); *P.s* (PSM), *Pisum sativum* mixed amyrin synthase (AB034803); *A.t* (LUP2), *Arabidopsis thaliana* multifunctional synthase (AF003472); *A.t* (LUP1), *Arabidopsis thaliana* lupeol synthase (U49919); *O.e* (OEW), *Olea europaea* lupeol synthase (AB025343); *T.o* (TRW), *Taraxacum officinale* lupeol synthase (AB025345). The phylogenetic tree was constructed by using the UPGMA method as implemented in the "Neighbor" program of the PHYLIP package (Version 3.5c) [50]. Amino acid distances were calculated using the Dayhoff PAM matrix method of the "Protdist" program of PHYLIP. The numbers indicate the numbers of bootstrap replications (out of 500) in which the given branching was observed. The protein parsimony method (the "Protpars" program of PHYLIP) produced trees with essentially identical topologies

Mechanism-based irreversible inhibitors and mutational analysis with OSCs have shown that the highly conserved amino acid motif DCTAE is required for substrate binding [27, 46, 57 – 59] (Table 3), and the conserved aspartate residue within this motif (D456) has been implicated as the likely electrophilic activator in the generation of the protosteryl cation for LS [57, 58, 60]. Similar experiments indicate that two aspartate residues at the homologous position of the

*Alicyclobacillus acidocaldarius* SC amino acid sequence (amino acids 376 and 377 of the DDTAV motif) are also essential for enzyme activity [59]. These residues are located in the large central cavity of the dimeric enzyme [55]. Interestingly, targeted mutations that convert the DDTAV motif of SC to DCTAE (the corresponding OSC motif) result in a change in substrate specificity from squalene to 2,3-oxidosqualene [61].

In addition to the DCTAE/DDTAV motifs, a highly conserved repetitive $\beta$-strand turn motif rich in aromatic amino acids (the QW motif) occurs in all OSCs and SCs and is repeated four to eight times (Table 3). These repeats are likely to be important for protein structure and stability and also for catalytic activity [55, 62–64]. The aromatic amino acids of the QW motif have been proposed to constitute sites of negative point charge that may interact with the intermediate cations during the cyclisation process [62].

### 2.3.2 *Processing and Post-Translational Modifications*

The predicted size for OSCs based on gene sequences is generally around 85 kDa. However, protein characterisation has indicated that the purified enzymes are often smaller than this. For example, the CS and $\beta$AS proteins are 55-kDa and 35-kDa respectively in pea [27, 30, 31], and 54-kDa and 28-kDa in *R. japonica* [32], while purified yeast LS has a molecular mass of 26 kDa [65]. These differences may be due to proteolysis associated with post-translational modification or to degradation during purification [38]. It is unlikely that the discrepancy between predicted and actual molecular mass is due to mRNA splicing, since Northern blot analysis of OSC gene expression normally gives a signal in the region of 2–3 kb, which is in the expected size range for the full length sequence [35, 38, 66]. There is evidence to indicate that OSCs are glycosylated, although incubation of rat LS with *N*-glycosidase did not affect activity [34].

### 2.3.3 *Product Determination*

Comparison of the amino acid sequences of OSCs that generate sterols and triterpenoids reveals a number of residues in addition to the DCTAE and QW motifs that are highly conserved in all classes. These residues may be required to open up the epoxide ring of 2,3-oxidosqualene, a catalytic step that is common to all of these enzymes [46]. Amino acid residues that are conserved exclusively in sterol OSCs or in triterpenoid OSCs may be required for formation of the protosteryl or dammarenyl cations, respectively.

Point mutations that give altered sterol profiles have been generated in *A. thaliana* and *S. cerevisiae* LSs. Matsuda and co-workers used a yeast expression system to select for spontaneous mutations in *A. thaliana* CS that restored sterol-independent growth to an LS-deficient mutant of yeast [67]. In this way they were able to identify a mutation from isoleucine to valine (at Ile481) that allowed synthesis of the sterols lanosterol and parkeol. Further studies have identified a number of other amino acid residues in *A. thaliana* CS and

**Table 3.** Conserved amino acid motifs of 2,3-oxidosqualene cyclases

| Organism | Enzyme | QW motif | QW motif | Substrate binding motif | QW motif | QW motif |
|---|---|---|---|---|---|---|
| *Alicyclobacillus acidocaldarius* | Squalene-hopene cyclase (SC) | ..RAVEYLLSCQKDEGYW.. | ..MEKIRRYLLHEQREDGTW.. | ..YPDVCDTAVVVWA.. | ..RRAVEYLKREQKPDGSW.. | ..KALDWVEQHQNPDGGW.. |
| *Rattus norvegicus* | Lanosterol synthase (LS) | ..NGVTFYAKLQAEDGHW.. | ..REEMVRYLRSVQLPNGGW.. | ..WIVADCTAEALKA.. | ..NQGLDFCRKKQRADGSW.. | ..QACHFLLSRQMADGGW.. |
| *Homo sapiens* | Lanosterol synthase (LS) | ..NGMTFYVGLQAEDGHW.. | ..REEIVRYLRSVQLPDGGW.. | ..WIVSDCTAEALKA.. | ..TQGLEFCRRQQRADGSW.. | ..RACDFLLSRQMADGGW.. |
| *Panax ginseng* | Cycloartenol synthase (PNX) | ..RAMSFYSTLQAHDGHW.. | ..KREICRYLYNHQNRDGGW.. | ..WPISDCTAEGFKA.. | ..EKAALFIEKIQSSDGSW.. | ..KACDFLLSKQVASGGW.. |
| *Panax ginseng* | $\beta$-Amyrin synthase (PNY) | ..RAVHFFSALQASDGHW.. | ..RKEILRYIYCHQNEDGGW.. | ..WQVSDCTAEGLKC.. | ..TNAVRYLEDTQMPDGSW.. | ..KAVEFLLKSQMDDGGW.. |
| *Taraxacum officinale* | Lupeol synthase (TRW) | ..RAISFYSTIQAHDGHW.. | ..QLEIKRYIYNHQNEDGGW.. | ..WQVSDCTAEGLKV.. | ..SRAVKYVEDTQESDGSW.. | ..KACKFLLSKQLPDGGW.. |

Database accession numbers and the regions of amino acid sequences shown for each of the five motifs are as follows: *Alicyclobacillus acidocaldarius* SC (AB007002): amino acids 17–32, 61–78, 372–384, 469–485, 518–533; *Rattus norvegicus* LS (U31352): amino acids 80–95, 125–142, 452–464, 562–578, 615–630; *Homo sapiens* LS (HS225261): amino acids 79–94, 124–141, 451–463, 561–577, 614–629; *Panax ginseng* PNX (AB009029); amino acids 99–114, 147–164, 479–491, 590–606, 640–655; *Panax ginseng* PNY (AB009030): amino acids 102–117, 150–167, 481–493, 593–609, 643–658; *Taraxacum officinale* TRW (AB025345): amino acids 100–115, 148–165, 481–493, 592–608, 642–657.

also in yeast LS that when mutated result in altered ratios of sterol products [68, 69].

The OSCs that give rise to triterpenoids are clearly distinct from LS and CS enzymes (Fig. 3). The LuS enzymes from *O. europaea* (OEW) and *T. officinale* (TRW) both produce lupeol as the sole product when expressed in yeast [45]. Despite their disparate taxonomic origins, these enzymes share 78 % amino acid identity [45, 70]. The ATLUP1 LuS from *A. thaliana* [46] is less closely related to these two enzymes, and differs from them in its deprotonation mechanism [15, 70]. ATLUP1 is also less specific in the products that it generates, and produces $\beta$-amyrin and other minor triterpenoids in addition to lupeol [46, 47]. This has led to the suggestion that two groups of LuS genes have arisen during the evolution of higher plants [45]. However the triterpene cyclase ATLUP2, which shares high amino acid sequence relatedness to ATLUP1 (Fig. 3), synthesises $\beta$-amyrin and $\alpha$-amyrin as major products in addition to lupeol [47, 48]. Thus the subgroup of OSCs that consists of ATLUP1 and ATLUP2 are multifunctional enzymes that are capable of generating a number of different triterpenoid products.

The *P. ginseng* $\beta$AS (PNY) and the *A. thaliana* LS (ATLUP1) cDNAs have been used to construct a series of chimeric proteins in order to investigate the regions of the enzymes that are important for product specificity [15]. The properties of the chimeric proteins were assessed following expression in yeast. These experiments identified an internal portion of the enzyme extending from approximately 25 – 50 % of the predicted amino acid sequence that was critical for $\beta$-amyrin formation. Interestingly, all of the chimeric enzymes were multifunctional, producing mainly $\beta$-amyrin and lupeol and also some other minor triterpenoids [15, 71]. Site-directed mutagenesis of the *P. ginseng* $\beta$AS PNY and the *O. europaea* LuS OEW has identified single amino acid residues that when mutated give interchanged product specificity [71]. Trp 259 of $\beta$AS is believed to control $\beta$-amyrin formation through stablization of the oleanyl cation, while the absence of cation stabilization with Leu in the equivalent position in LuS may result in termination of the reaction at the lupenyl cation stage. Mutation of a neighbouring tyrosine residue (Tyr261) that is conserved in all OSCs that produce pentacyclic triterpenoids to histidine, the corresponding residue in OSCs that produce tetracyclic carbon skeletons, resulted in synthesis of dammarenetype triterpenoids. Thus Tyr261 is proposed to play an important role in synthesis of pentacyclic triterpenoids by stabilising one of the cation intermediates generated after the dammarenyl cation [71].

It is clear that while some triterpenoid OSCs are highly specific in the products that they generate (such as the *P. ginseng* $\beta$AS enzymes [12, 14] and the LuS enzymes from *O. europaea* and *T. officinale* [45]), others are multifunctional and form a number of different products, at least in yeast [44, 46, 48, 49]. Although oleanane, lupane and dammarane skeletons are the most common triterpenoid structures associated with saponin biosynthesis, over 80 different oxidosqualene cyclase products are likely to be produced by plants [72]. The multifunctional properties of some triterpenoid OSCs and the fact that single amino acid changes can alter the nature of the product suggest that the structural diversity in nature may be generated by a core set of "flexible" enzymes. It is interesting to

note that, although DS enzymes have not yet been cloned from plants, mutations giving rise to a single amino acid change in other triterpenoid OSCs can confer the ability to synthesis dammarane-type triterpenoids [71].

### 2.4 Manipulation of Flux Through the Sterol and Triterpenoid Biosynthetic Pathways

Since OSCs are implicated in regulation of flux through the isoprenoid pathway leading to sterols and triterpenoids, manipulation of these enzyme activities may be expected to change the metabolic flow towards these two pathways [23, 24]. Several lines of evidence indicate that this is indeed the case. A number of studies have been carried out to investigate the effects of inhibitors of 2,3-oxidosqualene cyclases on sterol and triterpenoid synthesis. 8-Azadecalines bearing an isoprenoid-like N-substituent are excellent inhibitors of LS and CS, but do not inhibit $\beta$AS [73–76]. These inhibitors cause blocking of sterol biosynthesis at the level of CS, resulting in accumulation of $\beta$-amyrin at the expense of cycloartenol and methylene-24-cycloartenol.

Cell suspension cultures of *Gypsophila paniculata* and *Saponaria officinalis* produce very closely related triterpenoid saponins. Pretreatment of cell suspension cultures of *G. paniculata* with gypsogenin 3,*O*-glucuronide (a triterpenoid saponin precursor in *G. paniculata*) followed by administration of [$^{14}$C] acetate resulted in a marked reduction in incorporation of radioactivity into saponins and their precursors, but not into sterols and steryl glycosides [26]. Measurements of OSC activities revealed that there was no effect of elicitor treatment on CS levels in either species, but in *G. paniculata* $\beta$AS levels went down while in *S. officinalis* they increased. This suggests that in these two species OSCs are regulating steps in the isoprenoid pathway and control the flux to sterols and triterpenes.

In *Tabernaemontana divaricata* treatment of plant cell suspension cultures with an elicitor cause inhibition of CS activity [24, 25]. This response is accompanied by stimulation of activity of constitutive enzyme activities of the isoprenoid pathway leading to 2,3-oxidosqualene (squalene synthase and squalene oxidase), and induction of enzymes required for biosynthesis of pentacyclic triterpenoid phytoalexins ($\beta$AS and $\alpha$AS). Thus inhibition of the branchpoint enzyme CS results in increased flux through the triterpenoid pathway.

Now that a number of cloned CS, LuS and $\beta$AS genes are available it will be possible to generate transgenic plants in which the levels of these enzymes have been manipulated. The effects of altering the levels of these different OSCs on sterol and triterpenoid synthesis can then be assessed.

## 3 Elaboration of the Aglycone

The synthesis of saponins from the cyclisation product of 2,3-oxidosqualene involves a series of additional modifications. These may include a variety of oxidation and substitution events, and the addition of sugars at different positions on the skeleton [9]. Very little is known about the enzymes required for these

elaborations. However, one common feature shared by all saponins is the presence of a sugar chain attached to the aglycone at the C-3 hydroxyl position. The sugar chains differ substantially between saponins but are often branched, and may consist of up to five sugar molecules (usually glucose, arabinose, glucuronic acid, xylose or rhamnose) [9]. An understanding of the glycosylation process (which is believed to be the terminal stage in saponin biosynthesis) is important, since the presence of the C-3 sugar chain is critical for the biological activity of many saponins [2–11]. Substantial progress has been made in the characterisation of saponin glycosyl transferases, and evidence has emerged to indicate that the oligosaccharide chains are likely to be synthesised by sequential addition of single sugar molecules to the aglycone [77–79]. The majority of this work has involved steroidal glycoalkaloids and steroids, and there has been relatively little emphasis on triterpenoid glycosylation by comparison. However these studies are likely to have implications for analysis of triterpenoid glycosylation, and so merit inclusion in this review. The developments in this general area are summarised below.

### 3.1 Saponin Glycosyltransferases

All plants contain membrane-bound uridine 5′-diphosphate (UDP)-glucose: sterol glucosyltransferases that are required for synthesis of steroidal glycosides. Steroidal glycosides are common components of cell membranes and may play important roles in the regulation of membrane structures. In addition to these sterol glycosyltransferases, a number of plants have also been shown to have glycosyltransferase activities that catalyse the addition of monosaccharides to saponin aglycone substrates ("sapogenins") (Table 4). These enzymes differ from the sterol glycosyltransferases in that they are soluble, their activity is inhibited by Triton X-100, and they are more sensitive to heavy metals such as $Zn^{2+}$ and $Hg^{2+}$ than their membrane-bound counterparts [77, 80–94].

Two sugar transferases that glycosylate steroidal alkaloid aglycones have been purified to homogeneity. These are a UDP-galactose:tomatidine galactosyltransferase from tomato [93] and a UDP-glucose:solanidine glucosyltransferase

**Table 4.** Saponin glycosyltransferase activities in plants

| | Sapogenin | Sugar | Plant | References |
|---|---|---|---|---|
| Steroid | Nuatigenin | D-Glu | *Avena sativa* | 81–83 |
| | Sarsasapogenin | D-Glu | *Asparagus officinalis* | 84 |
| | Yamogenin | D-Glu | *Asparagus plumosus* | 85 |
| | Diosgenin | D-Glu | *Solanum melongena* | 77, 86, 87 |
| Steroidal alkaloid | Solasodine | D-Glu | *Solanum melongena* | 86, 87 |
| | Tomatidine | D-Gal | *Lycopersicon esculentum* | 91–93 |
| | Solanidine | D-Glu | *Solanum tuberosum* | 80, 88–90 |
| | Solanidine | D-Gal | *Solanum tuberosum* | 80, 89, 90 |
| Triterpenoid | Oleanolic acid | D-Glu | *Calendula officinalis* | 78, 79 |

from potato [88], which are believed to initiate sugar chain formation during the synthesis of the steroidal glycoalkaloids $\alpha$-tomatine and $\alpha$-solanine, respectively. The cDNA encoding the UDP-glucose:solanidine glucosyltransferase (StSGT) has been isolated by expression in yeast [95]. StSGT is the first saponin glycosyltransferase to be cloned from plants. The predicted amino acid sequence of this enzyme contains a conserved UDP-binding domain in common with other plant UDP-glucosyltransferases [95, 96], and also a domain that shares similarity with steroid-specific UDP-glucuronosyltransferases from mammals [95].

Clearly enzymes that are capable of catalysing the addition of single sugars to sapogenins are common in plants (Table 4), although the relatedness between these various enzymes will not become clear until more glycosyltransferases have been purified and/or the relevant genes cloned. The challenge will then be to characterise the other sugar transferase enzymes required for the subsequent glycosylation steps that give rise to the oligosaccharide chains that are typical of saponins. Although enzyme activities that transfer single sugars can be assayed in vitro, the possibility that in plants the oligosaccharide chains may be transferred intact to the aglycone should also be considered.

## 4 Conclusions

This review has addressed recent advances in two key areas of saponin biosynthesis, namely the cyclisation of 2,3-oxidosqualene and the glycosylation of sapogenins. Good progress has been made in cloning and characterising OSCs and in the elucidation of amino acids that are important for enzyme activity and product specificity; OSCs with altered properties have been generated by construction of chimeric enzymes [15], by targeted mutation of critical amino acid residues [27, 46, 57 – 64, 71] and also by the selection of spontaneous mutants in yeast [67]. Further advances in this area will give a comprehensive insight into the mechanistic enzymology and evolution of this fascinating superfamily of enzymes, and will enable the development of new enzymes that generate novel products by "accelerated evolution". Furthermore, the cloned sterol and triterpenoid OSCs can now be used to generate transgenic plants with altered levels of these cyclase activities (by overexpression or gene silencing). This will establish whether cyclisation of 2,3-oxidosqualene is a rate-limiting step for sterol and triterpenoid biosynthesis, and will enable the effects of manipulation of the different cyclase levels on the flux through these pathways to be investigated.

The synthesis of triterpenoid saponins from the skeletons shown in Fig. 2 involves a series of further modifications that may include a variety of different oxidation and substitution events [9]. Very little is known about the enzymes and genes involved in the elaboration of the triterpenoid skeleton, although genetic and biochemical analysis of saponin-deficient mutants of plants is likely to accelerate the dissection of these processes [16]. Progress has been made in the characterisation of saponin glucosyltransferases (primarily for steroidal and steroidal alkaloid saponins), and the first of these enzymes (StSGT from potato) has been cloned. Since glycosylation at the C-3 hydroxyl position confers am-

phipathic properties on the molecule and is normally critical for biological activity [2–11], this is clearly an important area in which to invest effort in the future.

**Acknowledgement.** The Sainsbury Laboratory is supported by the Gatbsy Charitable Foundation, and the support of Dupont is also gratefully acknowledged.

## References

1. Wink M (1999) Functions of plant secondary metabolites and their exploitation in biotechnology, Sheffield Academic Press, Sheffield
2. Tschesche R (1971) Advances in the chemistry of antibiotic substances from higher plants. In: Wagner H, Hörhammer L (eds) Pharmacognosy and phytochemistry. Springer, Berlin Heidelberg New York, p 274
3. Schönbeck F, Schlösser E (1976) Preformed substances as potential phytoprotectants. In: Heitefuss R, Williams PH (eds) Physiological plant pathology. Springer, Berlin Heidelberg New York, p 653
4. Osbourn AE (1996) Plant Cell 8:1821
5. Morrissey JP, Osbourn AE (1999) Microbiol Mol Biol Revs 63:708
6. Price KR, Johnson IT, Fenwick GR (1987) CRC Crit Rev Food Sci Nutr 26:27
7. Hostettmann K, Hostettmann M, Marston M (1991) Methods Plant Biochem 7:434
8. Fenwick GR, Price KR, Tsukamota C, Okubo K (1992) Saponins. In: D'Mello JP, Duffus CM, Duffus JH (eds) Toxic substances in crop plants. The Royal Society of Chemistry, Cambridge, UK, p 284
9. Hostettmann KA, Marston A (1995) Saponins. Chemistry and pharmacology of natural products. Cambridge University Press, Cambridge, UK
10. Roddick JG (1974) Phytochemistry 13:9
11. Osbourn AE (1996) Trends Plant Sci 1:4
12. Kushiro T, Shibuya M, Ebizuka Y (1998) Eur J Biochem 256:238
13. Kushiro T, Ohno Y, Shibuya M, Ebizuka Y (1997) Biol Pharm Bull 20:292
14. Kushiro T, Shibuya M, Ebizuka Y (1998) Molecular cloning of oxidosqualene cyclase cDNA from *Panax ginseng*: the isogene that encodes $\beta$-amyrin synthase. In: Ageta H, Aimi N, Ebizuka Y, Fujita T, Honda G (eds) Towards natural medicine research in the 21st century. Elsevier Science, Amsterdam, p 421
15. Kushiro T, Shibuya M, Ebizuka Y (1999) J Am Chem Soc 121:1208
16. Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE (1999) Proc Natl Acad Sci USA 96:12,923
17. Ruzicka L, Eschenmoser A, Heusser H (1953) Experientia 357
18. Eschenmoser A, Ruzicka L, Jeger O, Arigoni D (1955) Helv Chem Acta 38:1890
19. Nes WR, McKean ML (1977) Biochemistry of steroids and other isoprenoids. University Park Press, Baltimore
20. Abe I, Rohmer M, Prestwich GD (1993) Chem Rev 93:2189
21. Jolidon S, Polak AM, Guerry P, Hartmann PG (1990) Biochem Soc Trans 18:47
22. Cattel L, Ceruti M, Viola F, Delprino L, Balliano G, Duriatti A, Bouvier-Navé P (1986) Lipids 21:31
23. Baisted DJ (1971) Biochem J 124:375
24. Threlfall DR, Whitehead IM (1990) Redirection of terpenoid biosynthesis in elicitor-treated plant cell suspension cultures. In: Quinn PJ, Harwood JL (eds) Plant lipid biochemistry. Portland Press, London, p 344
25. Van der Deijden R, Threlfall DR, Verpoorte R, Whitehead IM (1989) Phytochemistry 28:2981
26. Henry M, Rahier A, Taton M (1992) Phytochemistry 31:3855
27. Abe I, Sankawa U, Ebizuka Y (1992) Chem Pharm Bull 40:1755

28. Goad LJ (1983) Biochem Soc Trans 548
29. Palmer MA, Bowden BN (1977) Phytochemistry 16:459
30. Abe I, Ebizuka Y, Sankawa U (1988) Chem Pharm Bull 36:5031
31. Abe I, Sankawa U, Ebizuka Y (1989) Chem Pharm Bull 37:536
32. Abe I, Ebizuka Y, Seo S, Sankawa U (1989) FEBS Letts 249:100
33. Buntel CJ, Griffin JH (1994) ACS Sym Ser 562:44
34. Abe I, Prestwich GD (1995) Proc Natl Acad Sci 92:9274
35. Kusano M, Shibuya M, Sankawa U, Ebizuka Y (1995) Biol Pharm Bull 18:195
36. Shi Z, Buntel CJ, Griffin JH (1994) Proc Natl Acad Sci 91:7370
37. Kelly R, Miller SM, Lai MH, Kirsch DR (1990) Gene 87:177
38. Roessner CA, Min C, Hardin SH, Harris-Haller LW, McCollum JC, Scott AL (1993) Gene 127:149
39. Corey EJ, Matsuda SPT, Bartel B (1993) Proc Natl Acad Sci 90:11,628
40. Sung CK, Shibuya M, Sankawa U, Ebizuka Y (1995) Biol Pharm Bull 18:1459
41. Corey EJ, Matsuda SPT, Bartel B (1994) Proc Natl Acad Sci USA 91:2211
42. Morita M, Shibuya M, Lee M-S, Sankawa U, Ebizuka Y (1997) Biol Pharm Bull 20:770
43. Venkatramesh M, Nes WD (1995) Arch Biochem Biophys 324:189
44. Morita M, Shibuya M, Kushiro T, Masuda K, Ebizuka Y (2000) Eur J Biochem 267:3453
45. Shibuya M, Zhang H, Endo A, Shishikura K, Kushiro T, Ebizuka Y (1999) Eur J Biochem 266:302
46. Herrera JBR, Bartel B, Wilson WK, Matsuda SPT (1998) Phytochemistry 49:1905
47. Husselstein-Muller T, Schaller H, Benveniste P (2001) Plant Mol Biol 45:63
48. Kushiro T, Shibuya M, Masuda K, Ebizuka Y (2000) Tetrahedron Letts 41:7705
49. Segura MJR, Meyer MM, Matsuda SPT (2000) Org Letts 2:2257
50. Felsenstein J (1996) Methods Enzymol 266:418
51. Ourisson G, Rohmer M, Poralla K (1987) Ann Rev Microbiol 41:301
52. Perzl M, Müller P, Poralla K, Kannenberg EL (1997) Microbiology 143:1235
53. Reipen IG, Poralla K, Sahm H, Sprenger GA (1995) Microbiology 141:155
54. Ochs D, Kaletta C, Entian K-D, Beck-Sickinger A, Poralla K (1992) J Bact 174:298
55. Wendt KU, Poralla K, Schultz GE (1997) Science 277:1811
56. Wendt KU, Lenhart A, Schulz GE (1999) J Mol Biol 286:175
57. Abe I, Prestwich GD (1994) J Biol Chem 269:802
58. Abe I, Prestwich GD (1995) Lipids 30:231
59. Feil C, Sussmuth R, Jung G, Poralla K (1996) Eur J Biochem 242:51
60. Corey EJ, Cheng H, Baker H, Matsuda SPT, Li D, Song X (1997) J Am Chem Soc 119:1277
61. Dang T, Prestwich GD (2000) Chem Biol 7:643
62. Poralla K, Hewelt A, Prestwich GD, Abe I, Reipen I, Sprenger G (1994) TIBS 19:157
63. Sato T, Kanai Y, Hoshino T (1998) Biosci Biotech Biochem 62:407
64. Poralla K (1994) Bioorg Med Chem Lett 4:285
65. Corey EJ, Matsuda SPT (1991) J Am Chem Soc 113:8172
66. Hayashi H, Hiraoka N, Ikeshiro Y, Kushiro T, Morita M, Shibuya M, Ebizuka Y (2000) Biol Pharm Bull 23:231
67. Hart EA, Hua L, Darr LB, Wilson WK, Pang J, Matsuda SPT (1999) J Am Chem Soc 121:9887
68. Herrera JBR, Wilson WK, Matsuda SPT (2000) J Am Chem Soc 122:6765
69. Meyer MM, Segura MJR, Wilson WK, Matsuda SPT (2000) Angew Chem Int Ed 39:4090
70. Kushiro T, Shibuya M, Ebizuka Y (1999) Tet Lett 40:5553
71. Kushiro T, Shibuya M, Masuda K, Ebizuka Y (2000) J Am Chem Soc 122:6816
72. Matsuda SPT (1998) In: Steinbüchel A (ed) Biochemical principles and mechanisms of biosynthesis and degradation of polymers. Wiley-VCH, Weinheim, p 300
73. Taton M, Benveniste P, Rahier A (1992) Biochemistry 31:7892
74. Taton M, Benveniste P, Rahier A (1986) Biochem Biophys Res Comm 138:764
75. Cattel L, Ceruti M (1992) 2,3-Oxidosqualene cyclase and squalene epoxidase: enzymology, mechanism and inhibitors. In: Patterson GW, Nes WD (eds) Physiology and biochemistry of sterols. American Oil Chemists' Society, Champaign, p 50
76. Taton M, Ceruti M, Cattel L, Rahier A (1996) Phytochemistry 43:75

77. Paczkowski C, Wojciechowski ZA (1994) Phytochemistry 35:1429
78. Kintia PK, Wojciechowski Z, Kasprzyk Z (1974) Bulletin de L'Académie Polonaise des Sciences 22:73
79. Wojciechowski ZA (1975) Phytochemistry 14:1749
80. Bergenstråhle A, Tillberg E, Jonsson L (1992) Plant Sci 84:35
81. Kalinowska M, Wojciechowski ZA (1986) Phytochemistry 25:2525
82. Kalinowska M, Wojciechowski ZA (1987) Phytochemistry 26:353
83. Kalinowska M, Wojciechowski ZA (1988) Plant Science 55:239
84. Paczkowski C, Wojciechowski ZA (1988) Phytochemistry 27:2743
85. Paczkowski C, Zimowski J, Krawczyk D, Wojciechowski ZA (1990) Phytochemistry 29:63
86. Paczkowski C, Kalinowska M, Wojciechowski ZA (1997) Acta Biochimica Polonica 44:43
87. Paczkowski C, Kalinowska M, Wojciechowski ZA (1998) Phytochemistry 48:1151
88. Stapleton A, Allen PV, Friedman M, Belknap WR (1991) J Agric Food Chem 39:1187
89. Zimowski J (1991) Phytochemistry 30:1827
90. Zimowski J (1992) Phytochemistry 31:2977
91. Zimowski J (1996) Enzymatic glycosylation of tomatidine in tomato plants. In: Waller GR, Yamasaki K (eds) Saponins used in traditional and modern medicine. Plenum Press, New York, p 71
92. Kalinowska M (1994) Phytochemistry 36:617
93. Zimowski J (1994) Acta Biochimica Polonica 41:202
94. Zimowski J (1998) Plant Sci 136:139
95. Moehs CP, Allen PV, Friedman M, Belknap WR (1997) Plant J 11:227
96. Vogt T, Jones P (2000) Trends Plant Sci 5:380

Received: June 2001