Evolution of Secondary Plant Metabolism

Secondary article

Peter G Waterman, *University of Strathclyde, Glasgow, UK*

The biosynthesis of secondary metabolites occurs universally in higher plants and shows very high structural diversity. Their evolution in higher plants rests on variation in the manipulation of a relatively small number of primary precursors.

Introduction

All higher plants have the capacity to produce secondary metabolites. The great majority of these metabolites originated from three precursor pathways, acetyl–coenzyme A (ketides), mevalonic acid (terpenes) and shikimic acid (aromatic amino acids and cinnamic acids). These pathways, singly and in combination, produce enormous structural diversity, with currently some 140 000 having been identified. This structural diversity is increased still further by widespread glycosylation and by the more occasional involvement of other primary metabolites, such as some nonaromatic amino acids and polysaccharides.

These compounds have been widely investigated by organic chemists because of their fascinating chemical diversity. More recently systematists have studied their distribution to aid the resolution of taxonomic problems and ecologists have become aware of the impact they can have through their interaction with the biotic and abiotic environment in which the producer exists. The supposition that secondary metabolites have evolved new structural forms during the evolution of higher plants is implicit in their use in taxonomy and their ecological role as defence against herbivores and pathogens and as cues for pollination and seed dispersal mechanisms (Harborne, 1993).

Chemosystematics of Plants

Chemically based characters (taste, smell, flower colour) have always played an important role in the phenetic identification of plants by humankind. However, the widespread investigation of secondary metabolites as a taxonomic tool did not begin until the 1950s with the first developments in chromatography as a technique that allowed their rapid separation and in spectrophotometric and spectroscopic techniques for their identification. A classic early example of this was the work of Bate-Smith (1962). He used herbarium material of several hundred specimens representative of most families of higher plants to investigate the distribution of eight compounds, two condensed tannin precursors, three flavonols, two simple cinnamic acids and ellagic acid (a hydrolysable tannin component). Among the findings made were that:

- condensed tannins were associated with woodiness;
- in herbaceous plants simple cinnamic acids showed greater methylation of phenolic groups than was the case in woody plants;
- hydrolysable tannins, the presence of which was associated with ellagic acid, were less widely distributed than condensed tannins; and
- the degree of phenolic substitution of the B-ring of flavonols appeared to be greatest in what are considered to be the most ancient plant families and to have reduced with evolutionary advancement.

Thus Bate-Smith interpreted secondary metabolites as either primitive or advanced, from an evolutionary standpoint, which is a critical conceptual development for chemosystematics.

A further important advance that took place in the 1960s was the considerable insight being gained into the biosynthetic origins of many secondary metabolites and the mechanisms by which structures were built up in plants. A classic example (Cordell, 1981) is the alkaloid quinine [I] which, while the final structure contains a quinoline heterocycle, was shown to originate from the amino acid tryptophan (**Figure 1**). Tryptophan normally gives rise to alkaloids possessing an indole or carboline heterocycle, such as vincoside [II]. By contrast, the furoquinoline alkaloids such as pteleine [III] originate from anthranilic acid and not from tryptophan. It was only through an understanding of the biosynthetic origins of compounds such as these alkaloids that their full systematic value could be recognized. While quinine and pteleine may share a common chemical structure in the quinoline nucleus, the biosynthetic origins of the former clearly demonstrate it to be more closely related to vincoside than to pteleine.

Most of the greatest successes of chemosystematics, in terms of influencing systems of plant classification, originated in the 1960s and 1970s (Harborne and Turner, 1984; Waterman and Gray, 1988). While much endeavour continues, it has, as the knowledge of the distribution of compounds has increased, become clear that individual compounds and classes of compounds are generally more

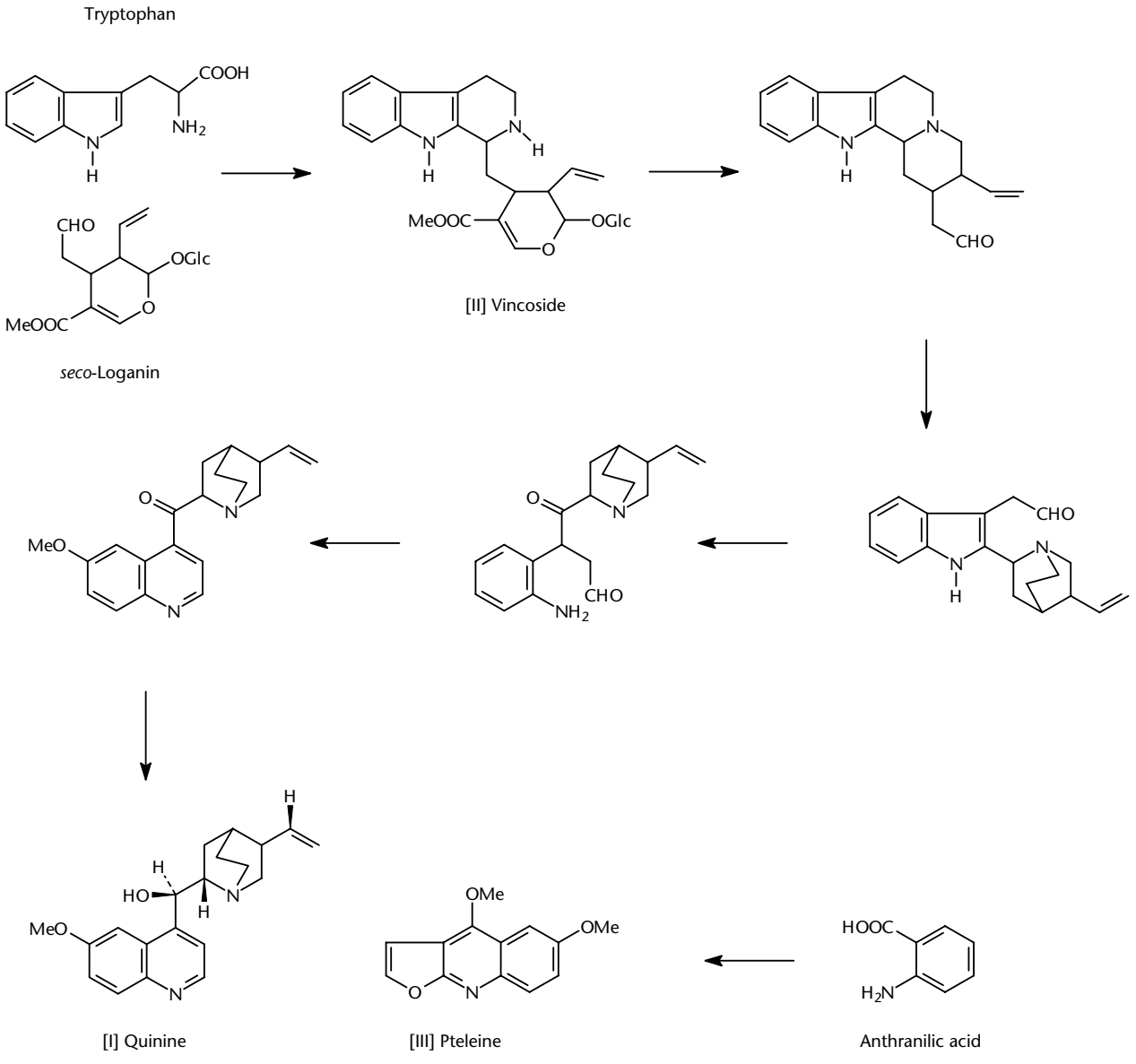


Figure 1 Origin of the quinoline nucleus of quinine [I] and pteleine [III] demonstrates convergent chemical evolution.

widely and often more sporadically distributed than was once thought and that this often confounds systematic usage. Why should this be? The old idea that when a compound ceased to be expressed it was because the power to express it had been lost is now not held to be universally true. In fact, the truly evolutionary event of either gain or loss of biosynthetic capacity through genomic change is a much rarer event than is genetically controlled suppression of biosynthesis that still, of course, leaves open the option for re-expression when evolutionary pressures change. Some examples of this are demonstrated in a comparison of DNA sequences and genetic and secondary metabolite expression data for the Fabaceae, where the expression of many classes of metabolites appears to be clearly polyphyletic while relatively few are monophyletic (Wink and Waterman, 1999). This situation is readily understandable if one accepts the argument that secondary metabolites are produced to interact with extrinsic factors to the overall benefit of the producer (Harborne, 1993).

Distribution Patterns of Metabolites

Despite the caveats explored in the last paragraph, it remains true that the distribution of secondary metabolites

is far from random. Some classes of metabolites are found throughout the entire realm of the higher plants, notably flavonoids, simple cinnamic acid derivatives and the common phytosterols such as β-sitosterol. However, if we take the flavonoids as an example, then within that general class we do see very pronounced distribution patterns that may in some cases relate to function. For example, condensed tannins, which are oligomers of flavonols, are associated with woodiness, anthocyanins and some flavonols are associated with flower colour, and classes of partially methoxylated aglycones are associated with leaf waxes. Isoflavonoids are found in greatest abundance in the Fabaceae, notably the Papilionoideae, while aurones are characteristic of the Asteraceae and the neoflavonoids occur mainly in genera within the Clusiaceae (Guttiferae) and in the tribe Dalbergieae of the Fabaceae, subfamily Papilionoideae.

In comparison with flavonoids, the distribution of alkaloids is limited, with only 20–30% of higher plants seemingly able to synthesize this major class of metabolite. A subdivision of the alkaloids with respect to their biosynthetic origin from a particular amino acid precursor (usually phenylalanine/tyrosine, tryptophan, ornithine, lysine, histidine or anthranilic acid) further restricts the distribution of each subclass. Consideration of the known distribution of the phenylalanine/tyrosine-derived alkaloids illustrates both the systematic value and the problems involved in using these compounds for chemosystematic purposes. In **Figure 2** the origins of five of the most important subtypes of alkaloids derived from these amino acids are recognized.

- Type A: 1-Benzyltetrahydroisoquinolines. Formed by the combination of tyrosine and a deaminated tyrosine in the form of a C_6–C_2–N–C_2–C_6 combination. The most widespread subtype of these alkaloids (see below for further discussion).
- Type B: *Erythrina* alkaloids. Formed by an unusual rearrangement of a type-A intermediate and found in the genus *Erythrina* (Fabaceae) together with some typical type-A alkaloids.
- Type C: Betalains. Formed from tyrosine and betalamic acid, which is itself a rearranged tyrosine unit. One of the few structural types of secondary metabolite that seem to be entirely restricted to one taxonomic order, the Caryophyllales (*sensu* Dahlgren, 1980), although they are not present in all families of that order (Mabry *et al*., 1978).
- Type D: Alkaloids of the Amaryllidaceae. Structurally related to type A but based on C_6–C_2–N–C_1–C_6 because the second amino acid-derived component has lost an additional carbon. Found in the Monocotyledonae, in some parts of the Liliaceae.
- Type E: Emetine-type alkaloids. A small subclass in which tyrosine has linked with a *seco*-loganic acid unit of mevalonate origin. Found only in a few species of the Rubiaceae.

The distribution of 1-benzyltetrahydroisoquinoline (1-btiq) alkaloids

The 1-btiq alkaloids (type A of **Figure 2**) have been of considerable taxonomic value. Their occurrence in the Papaveraceae and Fumariaceae was a critical factor supporting the recognition (Hegnauer, 1961) that these two families were part of what is now called the Annoniflorae (after Thorne, 1976) or Ranunculiflorae (after Dahlgren, 1980) rather than associated with nonalkaloidal families such as the Capparaceae. The distribution of 1-btiq alkaloids is centred on the Annoniflorae and allied Nymphaeiflorae (after Thorne, 1976), families that are responsible for more than 90% of all records of their occurrence (**Table 1**). However, as **Table 1** also illustrates, the small residue of records involves a further 11 families of very diverse taxonomic affinities, few among which have ever been taxonomically associated with the Annoniflorae and Nymphaeiflorae. In many of these other families, reports are limited and the diversity of structures isolated is low. However, in the Rutaceae (in 5 genera out of 140 studies) the complexity of the alkaloids produced matches that of the classical producing families (Waterman, 1998a), while in the Fabaceae (in *Erythrina*) they represent a separate subclass (type B).

The 'taxonomic anomalies' illustrated above for the distribution of 1-btiq alkaloids are normal for the distribution of most major classes of metabolites and could be illustrated over and over again by reference not only to alkaloids but to most other classes of secondary metabolite.

Evolutionary Trends in Secondary Metabolites

We are now well aware that the expression of biosynthetic potential can be an iterative process bringing into play genetic information that has remained silent in direct ancestors of the producer. We understand that evolutionary advancement may manifest itself in the generation of a new biosynthetic pathway or, just as likely, through the loss of a critical enzyme, giving a less complex sequence of reactions leading to a final product. We know that different biosynthetic processes can give rise to structurally similar compounds, so that the final structure in itself is not necessarily an indication of a close relationship. Given this degree of uncertainty, the prediction of evolutionary trends in the production of secondary metabolites remains a highly speculative process. The rest of the article will look briefly at possible evolutionary trends in a number of different classes of metabolites.

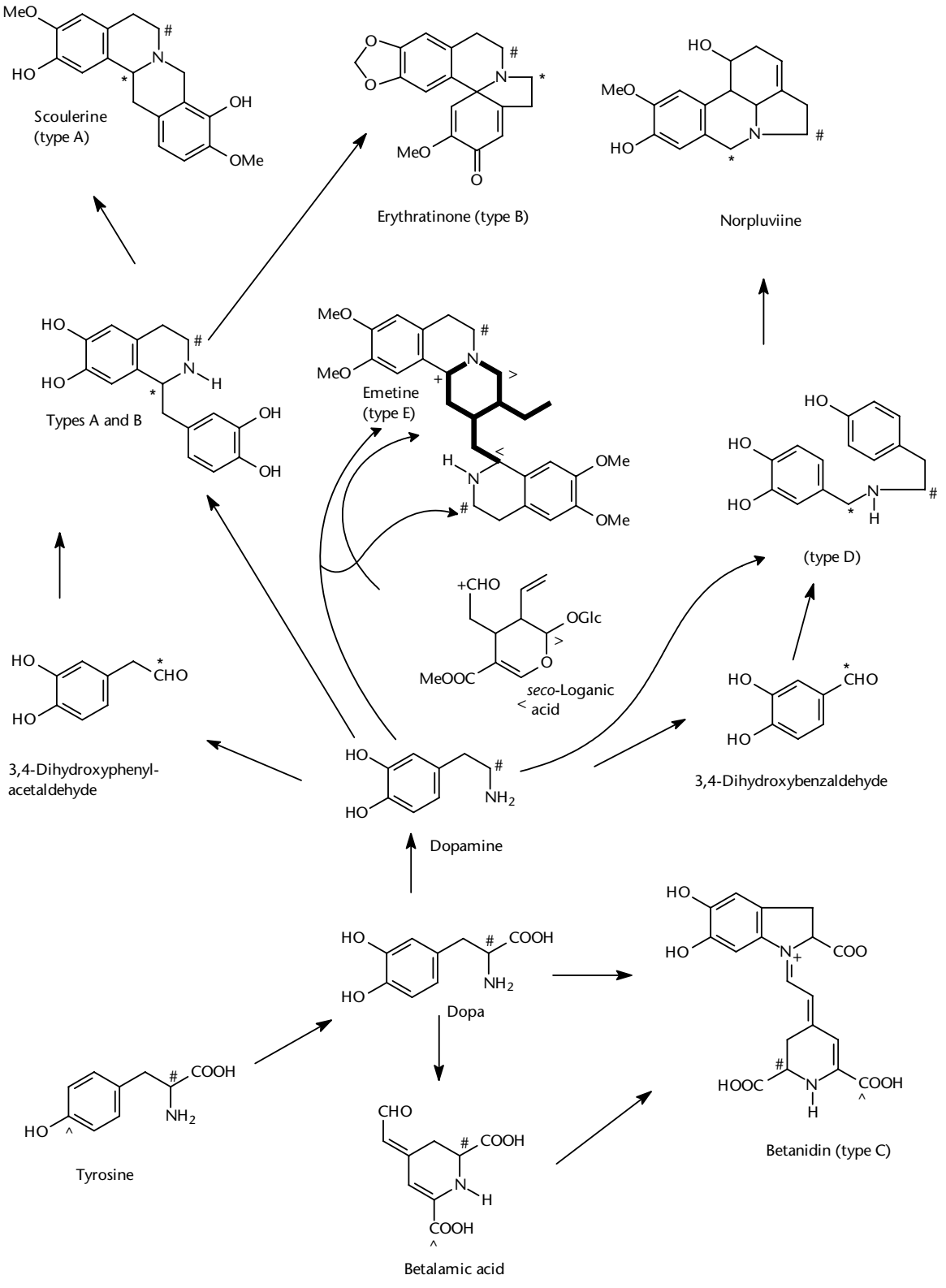


Figure 2 The evolution of five different classes of alkaloid from a common amino acid precursor, tyrosine. A given symbol always indicates the same carbon throughout the reaction scheme.

Table 1 Distribution of 1-benzyltetrahydroisoquinoline alkaloids in the Angiospermae

Families of Annoniflorae and Nymphaeiflorae	Other families
Annonaceae	Alangiaceae
Aristolochiaceae	Araceae (Monocot)
Berberidaceae	Buxaceae
Canellaceae	Caprifoliaceae
Eupomatiaceae	Euphorbiaceae
Fumariaceae	Fabaceae
Hernandiaceae	Liliaceae (Monocot)
Lauraceae	Rhamnaceae
Magnoliaceae	Rutaceae
Menispermaceae	Sapindaceae
Monimiaceae	Umbelliferae
Nympheaceae	
Papaveraceae	
Piperaceae	
Ranunculaceae	

Alkaloids

The major classes of alkaloid share an initial step in their biosynthesis. They all originate by the formation of a C–N bond (**Figure 3**), where the nitrogen is normally part of an amino acid and the carbon is from a ketone (usually an amide). The nitrogen and carbon may originate from different sources or as an 'internal' interaction where carbonyl and nitrogen occur in the same molecule, as in the formation of tropane alkaloids from ornithine (Waterman, 1998a). In synthetic organic chemistry terms, this reaction can be considered a classical Mannich condensation. Once that initial C–N bond is formed, then each alkaloid class undergoes structural proliferation.

The nature of alkaloid evolution seems therefore to be centred on arranging the juxtaposition between a nitrogen source and a carbonyl source that can then be joined together. The question arises of just what degree of genomic evolution is necessary to move from the production of one major alkaloid type to another (Waterman, 1998b). Given the similarities in the initiation of synthesis

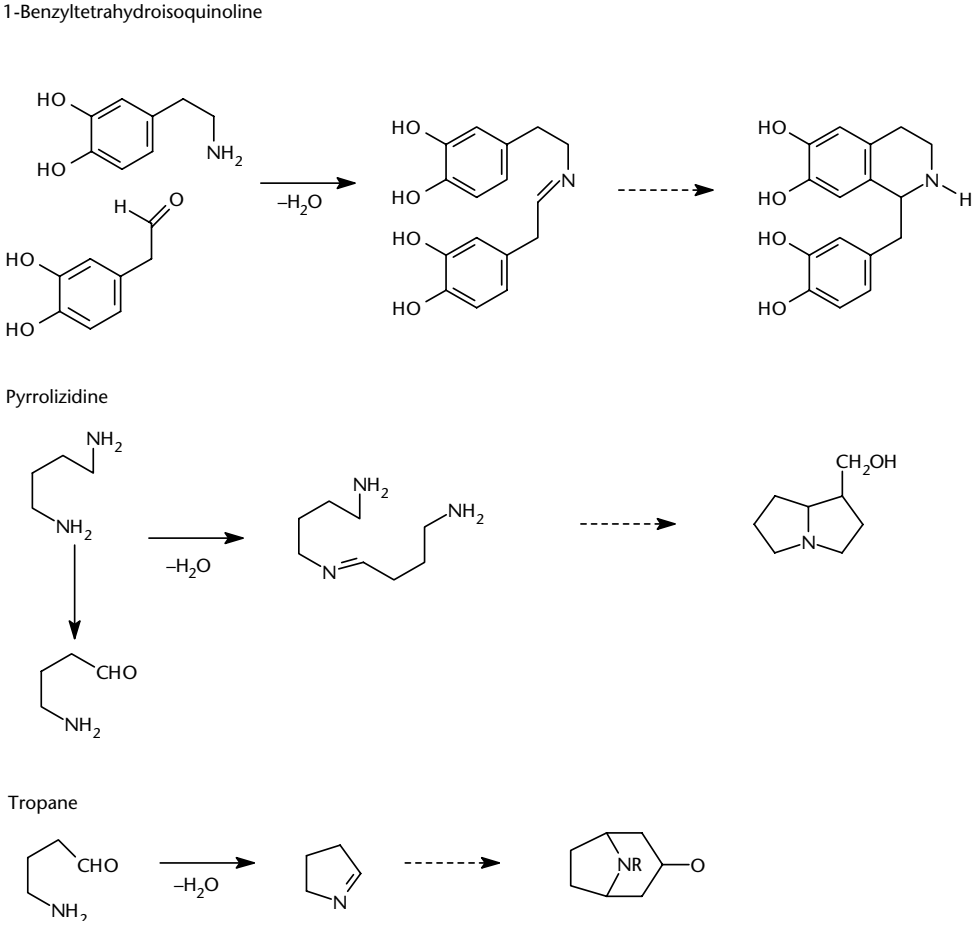


Figure 3 Formation of different classes of alkaloid employs a common strategy, suggesting that availability of a supply of precursors rather than unrelated *de novo* evolution of new biosynthetic pathways may be the ultimate requirement.

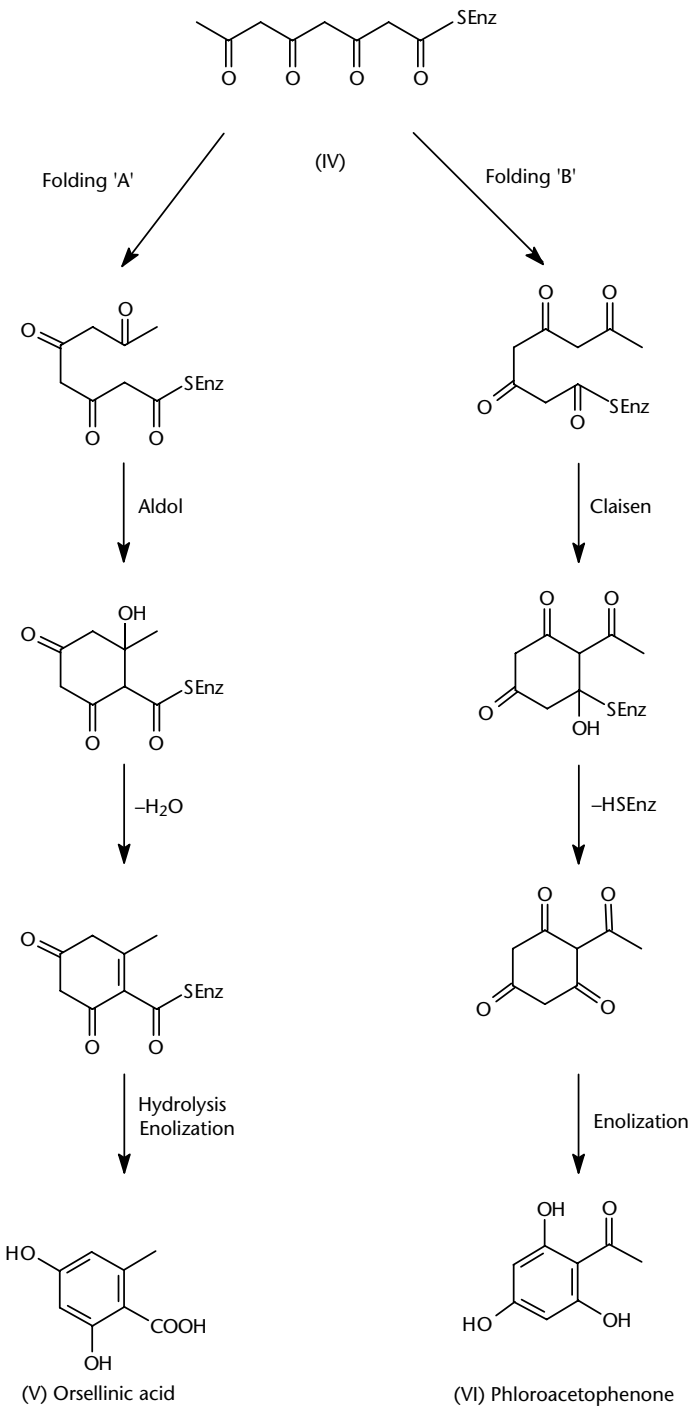


Figure 4 Different products can originate from the same precursor by changing the proximity of reactive groups within the precursor.

of all major alkaloid groups, how big a 'leap' is it to replace tyrosine with tryptophan or lysine as the nitrogen source or 3,4-dihydroxyphenylacetaldehyde with *seco*-loganin (see **Figure 2**) as the carbonyl source? We currently do not know the answer. Thus it is plausible that the key evolutionary event in the production of each of the major classes of alkaloids was the making available of surplus substrate that could be converted into these secondary metabolites to

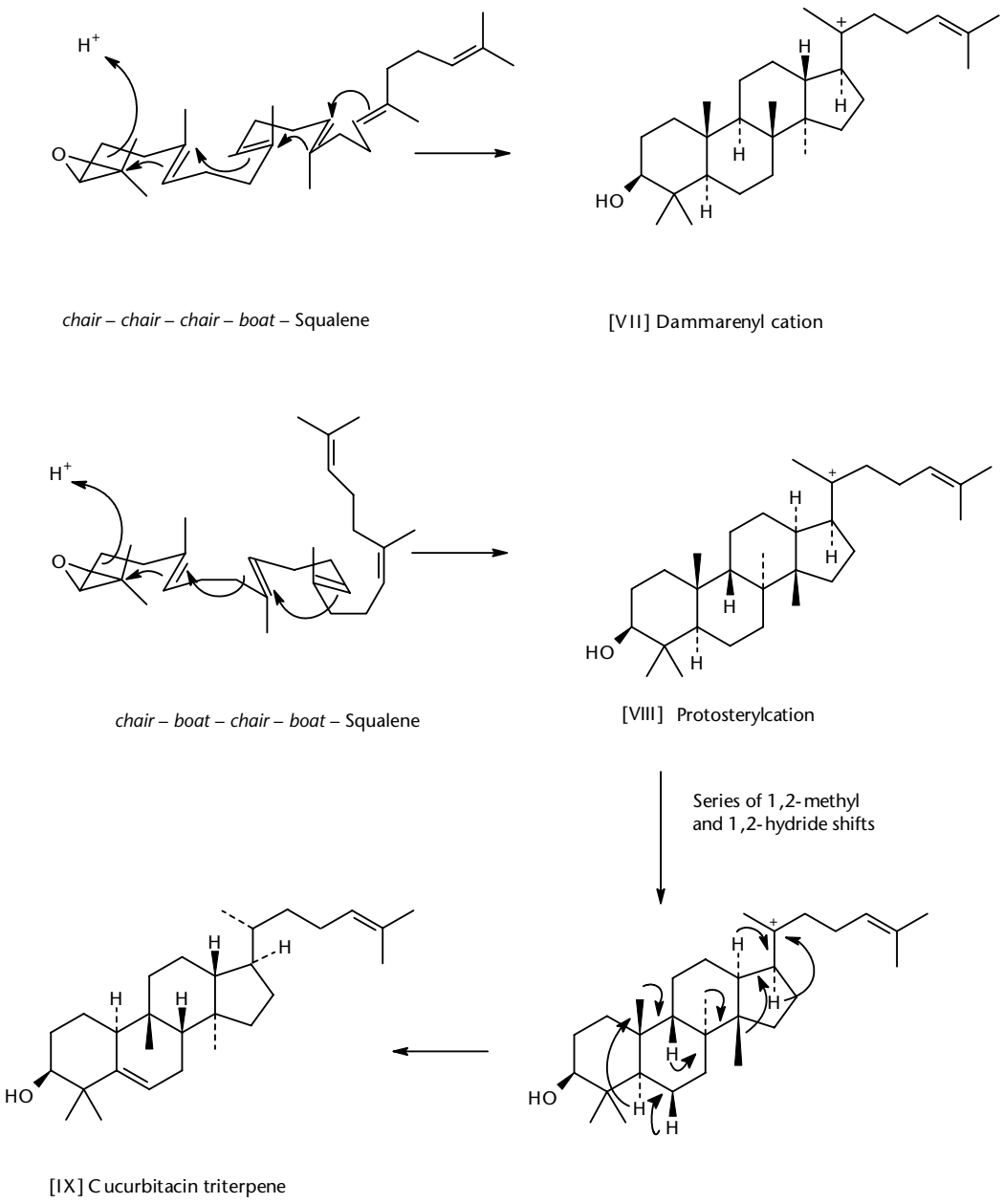
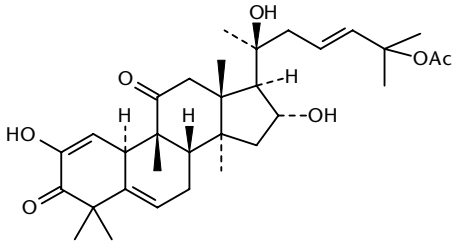


Figure 5 Variation in the stereochemistry of squalene epoxide leads to different intermediate triterpene cations.

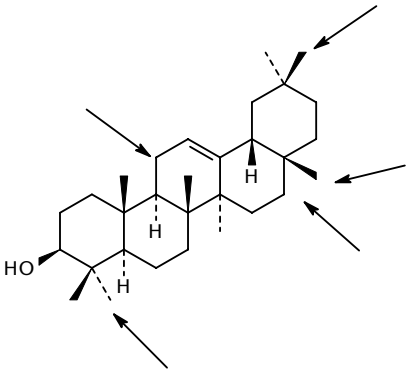
the net benefit of the producer. As pointed out elsewhere (Waterman, 1998b) the greatest evolutionary significance has always been associated with the generation of each alkaloid class rather than structural modification within that class. In terms of genomic evolution in alkaloids, we need to be a little cautious in making such interpretations.

Aromatic polyketides

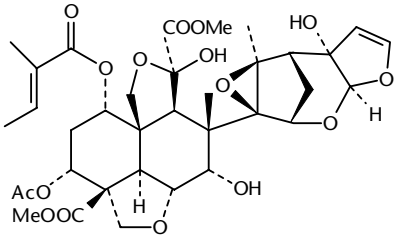
Acetyl–coenzyme A, through the mediation of malonyl–coenzyme A, undergoes polymerization reactions that give rise to a wide range of acyclic and cyclic products either as metabolites in their own right or as part of composite structures based on more than one of the primary



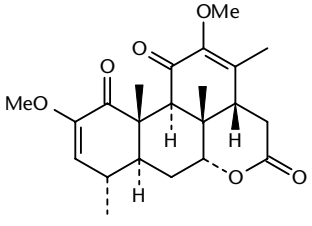
[X] Cucurbitacin-E



[XI] β -Amyrin, showing the carbons most commonly oxidized in saponins



[XII] Azadirachtin



[XIII] Quassin

Figure 6 The evolution of toxicity in triterpene derivatives has been through high levels of oxidation.

precursors (O'Hagan, 1993). The polyketides occur widely in both higher plants and fungi. In higher plants they are incorporated with a cinnamic acid precursor into flavonoids and so are ubiquitous.

As with alkaloids, evolutionary developments reflect the way the precursor is treated rather than profound differences in the starting materials. For example, the simple poly-β-keto ester [IV], which is generated by the condensation of four acetate units (**Figure 4**) can produce either a 6-methylsalicylic acid derivative such as orsellinic acid [V] or an acetophenone such as phloroacetophenone [VI]. It appears that the folding of the keto ester on the enzyme governs whether an aldol addition (in [V]) or a Claisen reaction (in [VI]) is favoured. In both cases the production is initiated by bringing into proximity a keto group and an activated methylene, indicating a conservatism in the initial stage of secondary metabolism similar to that seen in the alkaloids. In this case the 6-methylsalicylic pathway appears to have evolved in fungi, whereas the acetophenone route is generally favoured in higher plants. It should be noted that salicylic acid, found in a number of higher plants, is a product of the shikimic acid pathway and is not biosynthetically allied to orsellinic acid.

Triterpenes

The mevalonate pathway, like the acetate pathway, commences with a polymerization, this time of a 5-carbon unit derived from mevalonic acid. A huge array of metabolites arise from oligomers made up of two, three, four, six (triterpenes) or eight of these units. In addition, the monomer is a very active alkylating agent that finds its way into many essentially non-mevalonate metabolites, where it can often cyclize to give pyran or furan rings. Aberrant combinations of monomers occur with some frequency, leading to unusual classes of compounds such as the pyrethroids and iridoids (including *seco*-loganin, see **Figure 2**), but the acyclic hexamer squalene that leads to the

ubiquitous triterpenes represents a highly conserved pathway.

The cyclization of squalene (**Figure 5**) generally occurs through the 2,3-epoxide, protonation of which initiates a sequence of events leading to a tetracyclic ion. Depending on the conformation of the squalene, different ions can be generated such as the dammarenyl cation [VII] or the protosteryl cation [VIII]. The charge can be removed from this cation by (a) quenching with a hydroxyl; or (b) a series of internal methyl shifts (as in the cucurbitacins, [IX], derived from the prosteryl ion) followed by quenching through either addition of hydroxyl or loss of hydrogen to form a double bond; or (c) expansion of the D-ring and formation of a fifth ring to give the very common lupane, oleanane and ursane triterpene skeletons, once more terminating with quenching of the charge.

The pentacyclic triterpenes are ubiquitous in higher plants, but some of the tetracyclic types, notably the cucurbitacins, are far more unusual and have a distribution associated with a relatively small number of families.

A common feature among these triterpenes is the relative paucity of oxidation, which is restricted to C3 (from the epoxide) and to one other carbon when the cation is quenched by hydroxylation. Given that this is the 'starting position' for triterpenes, it is evident that oxidation is a key evolutionary development in many families (**Figure 6**). Such oxidation invariably imparts biological activity into the terpene that is generally beneficial in protecting against predators and/or pathogens.

For example cucurbitacin-E [X], which is a typical highly toxic cucurbitacin, shows additional oxidation at C2, C11, C16 and C20 in comparison with the initial cucurbitacin nucleus (see **Figure 5**). The saponins, a series of oxygenated triterpene glycosides deriving most notably from oleanane or ursane pentacyclic triterpenes such as β-amyrin, are often oxygenated at C11, C16, C21, C22 and at several of the methyl substituents [XI]. However, perhaps the most extreme example of oxidation of the triterpene occurs in the families of the Rutales (Rutaceae, Meliaceae, Simaroubaceae, Cneoraceae). Here the tirucallane triterpene nucleus has been heavily oxidized and rearranged (Waterman, 1993) to produce the limonoids and quassinoids with highly modified structures such as azadirachtin [XII] and quassin [XIII]. The limonoids and quassinoids are currently confined to the Rutales, suggesting that their formation may be one of the few truly unique evolutionary events in higher plant metabolism.

References

Bate-Smith EC (1962) The phenolic constituents of plants and their taxonomic significance. I. *Journal of the Linnean Society (Botany)* **58**: 95–173.

Cordell GA (1981) *Introduction to Alkaloids: A Biogenetic Approach*. New York: Wiley.

Dahlgren RMT (1980) A revised system of classification of the Angiospermae. *Botanical Journal of the Linnean Society* **90**: 91–124.

Harborne JB (1993) *Introduction to Ecological Biochemistry*, 4th edn. London: Academic Press.

Harborne JB and Turner BL (1984) *Plant Chemosystematics*. London: Academic Press.

Hegnauer R (1961) Gliederung der Rhoedales *sensu* Wettstein in lichte der inhaltstoffe. *Planta Medica* **9**: 37–46.

Mabry TG, Neuman P and Philipson WR (1978) *Hectorella*, a member of the betalain sub-order Chenopodiineae of the order Centrospermae. *Plant Systematics and Evolution* **130**: 163–165.

O'Hagan D (1993) *Polyketides*. Chichester: Ellis Horwood.

Thorne RF (1976) A phylogenetic classification of the Angiospermae. *Evolutionary Biology* **9**: 35–106.

Waterman PG (1993) Phytochemical diversity in the Order Rutales. In: Downum KR, Romeo JT and Stafford HA (eds) *Phytochemical Potential of Tropical Plants,* pp. 203–233. New York: Plenum.

Waterman PG (1998a) Chemical taxonomy of alkaloids. In: Roberts MF and Wink M (eds) *Alkaloids: Biochemistry, Ecology and Medicinal Applications,* pp. 87–107. London: Plenum Press.

Waterman PG (1998b) Alkaloid chemosystematics. In: Cordell GA (ed.) *The Alkaloids,* vol. 50, pp. 537–565. San Diego: Academic Press.

Waterman PG and Gray AI (1988) Chemical systematics. *Natural Products Reports* **4**: 175–203.

Wink M and Waterman PG (1999) Chemotaxonomy in relation to molecular phylogeny of plants. In: Wink M (ed.) *The Role of Plant Secondary Metabolites and Their Utilisation by Biotechnology*, pp. 299–340. Sheffield: Sheffield Academic Press.