

# BIOSYNTHETIC PATHWAYS IN CHEMICAL PHYLOGENY

ARTHUR J. BIRCH

*Research School of Chemistry, Australian National University,  
Canberra*

## ABSTRACT

An approach to chemical phylogeny is discussed, based on an emphasis on alterations of biosynthetic pathways rather than on isolations of individual compounds. Some offshoots of biosynthetic pathways to target molecules such as gibberellic acid are considered. Attempts are made to assess the significance of alterations of pathways in terms of the chemical probabilities of their occurrences.

---

Attempts have been made to assist that essentially artificial classification, taxonomy, by considering the structures of plant constituents as markers on a level with morphological characteristics. This is quite legitimate, but both taxonomy and its much more fundamental relative phylogeny may well be illuminated much more by considerations of biosynthetic pathways than of individual structures. Some tentative attempts have been made<sup>1</sup> but there appears to be required a more systematic analysis of the whole situation.

The constituents of plants and microorganisms can be divided roughly into two categories; molecules which are of primary metabolic and structural importance, and molecules of secondary, or no known, functions. The first class contains small molecules which appear to be ubiquitous, such as sugars, amino acids, citric acid, etc., which are the universal building blocks and energy sources. It also contains proteins, nucleic acids, polysaccharides, etc., which may differ in structural details from one organism to another but which appear to have universal *types* of function as enzymes, structural materials and carriers of hereditary mechanisms. More arguable as to category are rather specialized structural materials characteristic of classes of whole organisms rather than of individual cells, such as lignin. The secondary category includes a great diversity of chemical types, usually in fairly small molecules, with a sporadic occurrence.

Although the first class is a more fundamental one, and undoubtedly will be examined in structural terms to define relations between organisms in a way which has already been done, for example, with the globin portion of haemoglobins<sup>2</sup>, the task is much more difficult because of the great molecular complexity involved and because of the universal distribution of types in which alterations may be small, for example, the replacement of one amino acid by another in a molecule containing many such units.

The secondary products, because they seem to mark diversity rather than similarity, are therefore obvious targets for examination in connection with possible evolutionary sequences. They are often experimentally simpler to deal with.

Many of the secondary metabolites do seem to have definable types of function, such as the pigmentations or odours of flowers which serve to attract fertilizing insects, or protective agents which act against predators or which help the organism to compete with others. Unlike the first class, where often very exact chemical structures seem to be required for specific biological activities, similar objectives are often achieved by secondary metabolites in a variety of chemically unrelated ways. Protective coatings on leaves, for example, represent a desirable set of physical properties, which can be attained by different plants using a number of chemical types: aliphatic hydrocarbons or their functional derivatives of differing chain-lengths and differing functionalities and chemically unrelated and diverse compounds such as triterpenes and their various functional derivatives.

Such metabolites can in fact be related frequently to a target activity, and in the chemical evolutionary sense sometimes to a target molecule to which they are related.

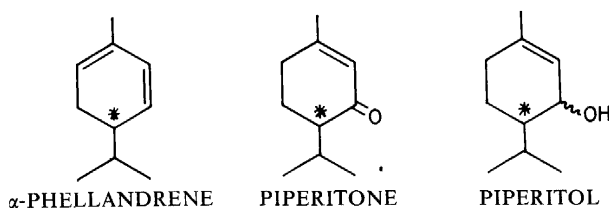
A good example of the first is pigmentation. It is of no consequence what chemical type of pigment is present in a flower or bract, provided it does the right job; carotenoids, flavonoids, anthocyanins and betacyanins alone or in combination can often provide similar results. They all consist of molecules which can chemically be readily modified: carotenoids by alteration of conjugated unsaturation and insertion of oxygenated groups, flavonoids and anthocyanins by alteration of oxygenation patterns, by attachment of glycoside and other groups and by complexing. Betacyanins can be altered by involvement of different amino acids in basically the same kind of synthetic process. A large number of different but related substances, in different combinations, is encountered therefore.

A target molecule is one which is required for some genuinely metabolic reason and therefore a biosynthetic process exists for producing it. Examples of two rather different types are the diterpenoid gibberellic acids, which control some plant metabolic processes, and farnesyl pyrophosphate, which is an intermediate required for polyterpene, including gibberellic acid, and steroid synthesis. The results of mutations which alter the biosynthetic processes to such molecules may be to form related substances instead, if they have the required biological activity, or in addition. Gibberellic acids, for example, occur in a considerable number of chemically slightly different, but biologically active forms, many related plant diterpenes which are not active as hormones may be useful in other ways, for example, as protective agents, and are clearly offshoots of the same biosynthetic sequence. Because of possible gene duplications, there is no problem concerning the production of the original hormone as well as non-hormonal congeners. Lower plant terpenes can all be related to the route which produces farnesyl pyrophosphate and its precursors.

'Physiological forms' of even the same species, as defined morphologically, are known which differ in their development of such secondary pathways. For example, a microorganism may be found in forms which do or do not

produce a given antibiotic, and trees are known which differ in secondary constituents, such as *Eucalyptus dives*, of which three forms exist with the leaf oils containing chiefly: (i) piperitone; (ii)  $\alpha$ -phellandrene; or (iii) cineol. The first two have the same absolute configuration at the asymmetric centre (\*) and the hydrocarbon is probably derived by dehydration of the piperitol shown (*Figure 1*) after allylic rearrangement. Cineol is probably related only through the common precursor geranyl pyrophosphate.

We tentatively assume that mutations of routes to a target molecule can occur and provided the target molecule itself or some biochemical equivalent is still formed that the species will survive for the periods in evolutionary time on condition that the resulting mutant molecules are useful, or at any rate not harmful.



*Figure 1*

A secondary metabolite is elaborated in a series of stages from a primary metabolic pool of small molecules, and the stages require enzymes, which are determined genetically, although conditions, such as the presence of a substrate, may be required to trigger the enzyme production. The amount of enzyme present may also be determined by the multiplicity of a gene, if polyploid. A recessive form of a gene presumably forms no active enzyme. An inhibitor gene may prevent expression in terms of enzyme activity. Conversely a gene may be required to unmask another leading to production of an enzyme system. Whatever the detailed genetic mechanism, the important factor is that transforming enzymes are genetically controlled.

Mutations of a gene may destroy enzyme activity entirely, the more probable process, or rather rarely alter its structure specificity. Breackage of a sequence by loss of an enzyme does not necessarily result in accumulation of the substance immediately before the break. Owing to feedback mechanisms, competing processes, and sometimes reversibility of the route, a precursor earlier still in the sequence, some rather unrelated product due to diversion of a basic building block, or no recognizably related product at all, may result. In some cases, such as the desmethyltetracyclines, a minor structural alteration of an intermediate may still permit completion of a sequence to a somewhat altered endproduct; in other cases where enzymic specificity is absolutely linked to a given structure, no further steps are possible after even a minor change. It is possible, as pointed out by Bu'lock<sup>3</sup>, that, with some enzymic elasticity, 'metabolic grids' may exist where basically similar steps may be possible in alternative sequences.

Breakage of a route may make available substances which, if fairly fundamental, e.g. acetyl coenzyme-A from arrest of a polyketide, may be shunted into a completely different route which hitherto has competed badly and may not even have been manifest. If a more specialized substance such as a diterpene results, it may be accumulated, or modified by the development of other processes, such as oxidation or cyclization, in ways not associated with the original route. Apparently new compounds may in any case represent merely inadequacies of analytical methods in detecting small quantities of minor products in an organism before mutation occurs. A probable example of the first type is the modification of *Penicillium brevicompactum* on continued culture, from the production of mycophenolic acid (a polyketide with introduced methionine Me and a terpene chain) to a series of C<sub>10</sub>-phenols (solely polyketide)<sup>4</sup>. Speculatively the alteration is due to loss of a step, such as C-methylation in the mycophenolic acid synthesis, and diversion of acetyl coenzyme-A into the other polyketide synthesis giving as major products what were initially minor ones (Figure 2).

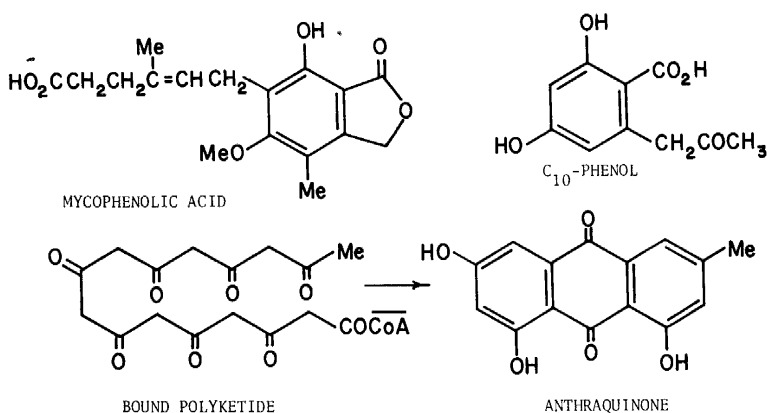


Figure 2

In a related case an all-or-nothing effect is observed. Mutations of *Penicillium islandicum* which normally gives polyketide anthraquinones, result either in anthraquinone or no recognizable polyketide product<sup>5</sup>. Presumably an enzyme-bound polyketide chain (Figure 2) merely 'unzips' to give acetylmalonyl coenzyme-A (which can enter the general metabolic pool) unless some chemically irreversible structure (in this case probably an aromatic ring) is reached in the sequence. It is interesting in this case that compounds with less than three rings are not produced which may indicate that the first ring-closure is a peripheral one (14-membered ring) six-membered ring folding then occurring rather automatically.

An acute question which arises in trying to assess the significance of a particular process to the relationship between two organisms is how likely

is such a process to be developed independently? In view of the limited number of precursors available from general metabolism and the limited number of mechanistic types of biochemical reactions the same routes could be developed in different organisms independently to give rise finally to the same or closely related products. Conversely the same molecular structure may arise on occasion by entirely different biochemical routes, so the emphasis is, as has been noted, on the *route* and not on the *product*.

The significance of the alteration of a process is clearly related to its degree of 'improbability'. This is a difficult term to define, and the problem is not a purely chemical one, since, provided a reaction is mechanistically possible, there seems no reason why an enzyme should not be capable of development to catalyse it. However, by a general survey of the situation it seems a fairly common rule that the more readily a reaction occurs chemically, the more likely an enzyme is to be produced to catalyse it. Most chemical theories of biosynthesis, many of which have been confirmed in considerable detail, e.g. refs 3, 6, are implicitly based on this assumption.

Independent developments of the same biosynthetic reaction in different organisms may well therefore be more frequent, the greater their ease as recognized by the organic chemist. Conversely the chemist has to recognize from the frequent occurrence of certain mechanistically recognizable reactions that they are biochemically probable. Modification of basic structures, by reactions such as ring-closures, O-methylations, etc., are often relatively 'probable' processes as viewed by the organic chemist from laboratory experience. Considerations are assisted by the known frequency of occurrences in Nature, but caution must be exercised. Natural acetylenes, for example, were once thought to be very rare; they turn out to be quite common once methods were devised to look for them, and the direct oxidation of a double to a triple bond seems a simple process in Nature compared to that in the laboratory. Some compounds may also be artifacts of isolation, a point always to be borne in mind.

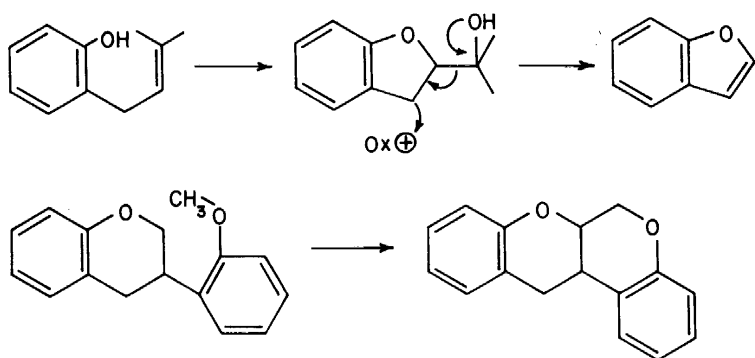
One notable area where laboratory experience may be misleading is that of specific oxidations at unactivated saturated positions of a molecule in ways which the chemist would consider highly improbable, but now has to accept as a common occurrence.

C-Methylation of natural molecules from methionine is a mechanistically acceptable process, and was predicted on this basis<sup>7</sup>. However, occurrences of C-methylated compounds are far fewer than of O- or N-methylated ones. Developments of enzymes for C-methylation processes are therefore rarer than for O- or N-methylations, as the chemist would expect, and its occurrence would be considered therefore as much more significant. The introductions of terpene units by alkylation on carbon is much commoner, compared with O-alkylation, than the chemist would expect, since polyterpene cations (from the pyrophosphate) would, like the methyl cations from methionine, be expected to O-alkylate readily. However, since the biochemically fundamental process of terpene elaboration requires alkylation on carbon, other processes, such as alkylation of aromatic rings, may have developed by modifications of the terpenoid C-alkylation enzymes. Terpene ethers do, of course, occur.

Considerations of both mechanistic probability and known natural

occurrence must therefore necessarily be involved in assessments of probabilities of reactions.

In considering the significance of a given compound its known or deducible place in a sequence must be assessed. For example, a methylenedioxybenzene derivative is a more 'advanced' compound than a methoxyhydroxybenzene from which it is oxidatively derived, as is a rotenoid with an extra pyran ring compared with a 2'-methoxyisoflavonoid, the pyran ring being derived by oxidative cyclization of the OMe (*Figure 3*). A natural benzofuran is subsequent to an orthoisopentenyl phenol, from which it is derived by loss of three carbon atoms, despite the fact that it is a simpler compound chemically (*Figure 3*), the more probable mutation which would occur with loss of a biosynthetic stage, would therefore in this case cause the products to appear to be more complex.



*Figure 3*

The assumption is therefore not even justified that a plant with the more advanced biogenetic sequence is necessarily the latest in evolution; related plants may have regressed as the result of probable mutations which abolish transformations. Partly for this reason consideration of only one type of synthesis may not be very helpful in evolutionary terms, and it is necessary to consider as many unrelated sequences as possible.

Sometimes chemical considerations can be invoked to assist in elucidation of relationships. *Cinnamomum* oils consist principally of three  $C_6-C_3$  types: containing mainly eugenol, mainly safrole and mainly cinnamyl alcohol-cinnamic aldehyde. The relationship between the first two is clear: eugenol  $\rightarrow$  safrole (OMe, OH  $\rightarrow$  OCH<sub>2</sub>O) (*Figure 4*) although whether in comparing two plants, one with eugenol and one with safrole, mutation has occurred with loss of a stage from safrole or gain of one from eugenol cannot be asserted on this evidence alone. The relation between the related cinnamyl alcohol-cinnamic aldehyde pair (with no oxygen in the ring and oxygen in the side-chain) and the eugenol-safrole pair (oxygen pattern reversed) is that, from

the probable mechanism of the process via a carbonium ion<sup>8</sup> (Figure 4), removal by reduction of OH from the sidechain of cinnamyl alcohols (possibly as phosphate esters) occurs readily only when *p*-oxygen is present. For whatever biochemical reason, para-oxygenation is less extensive in plants giving the cinnamyl alcohol series, and the sidechain deoxygenated compounds are in lower proportions. Interpretation is again doubtful, whether the cinnamyl alcohol plants have partially lost the para-hydroxylation ability, or whether the eugenol plants have gained it. The former seems more likely since the phenols are required for lignan synthesis.

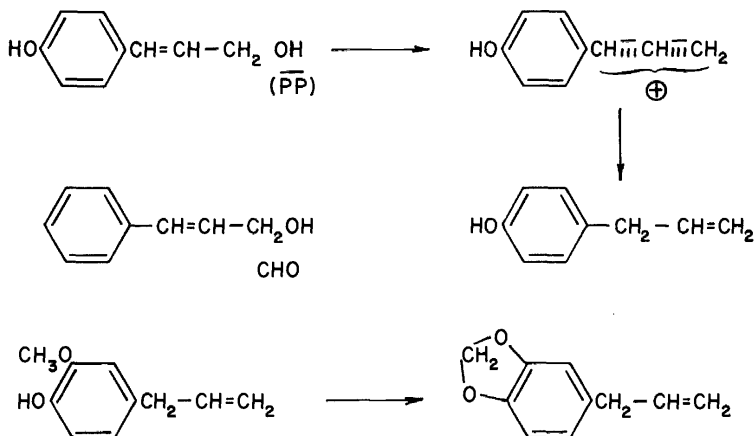
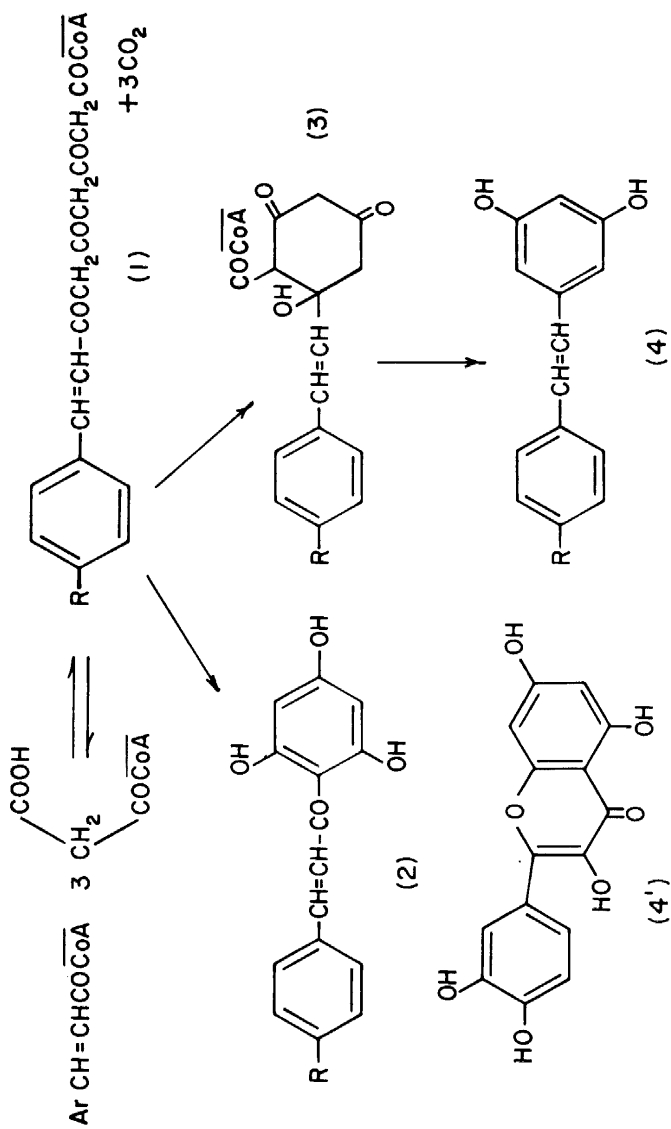


Figure 4

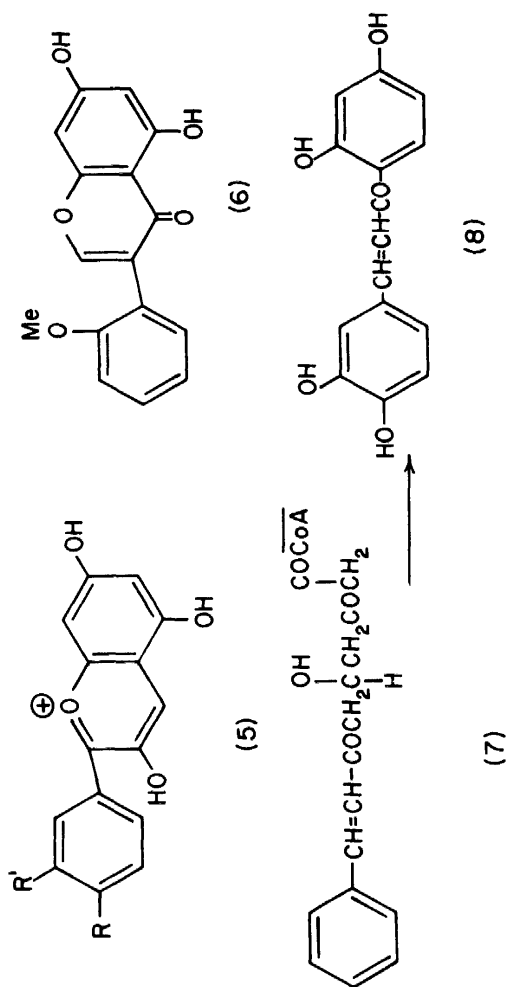
In order to illuminate the effect of genetics on a series of secondary metabolites, albeit of only marginal importance in phylogeny, let us consider one of the few cases where some information is available: the flower pigments, and in particular those of *Dahlia*. The chemical reactions involved in this case are chiefly oxidations and reductions.

Most of the genetic work was unfortunately done before modern methods of analysis and before the biosynthetic background was known. It has been possible to interpret<sup>9</sup> at least partially the genetic evidence once the origin of the pigments in a C<sub>6</sub>C<sub>3</sub> plus polyketide acetate route was suspected and proved.

Firstly, let us examine briefly the origins of flavonoids and related substances. The cinnamyl polyketide (1) (R may be H or OH) is an enzyme-bound and probably reversible precursor to the chalcone (2). An alternative ring-closure via (3) to (4) explains<sup>10</sup> the production of the plant stilbenes which occur with flavanones in pines and as an alternative in some *Eucalyptus* species. The difference between the flavonoids and stilbenes despite appearances is therefore relatively small and due chiefly to one ring-closure alteration in a common C<sub>15</sub>-monocyclic precursor. Apart from ring-closure of a

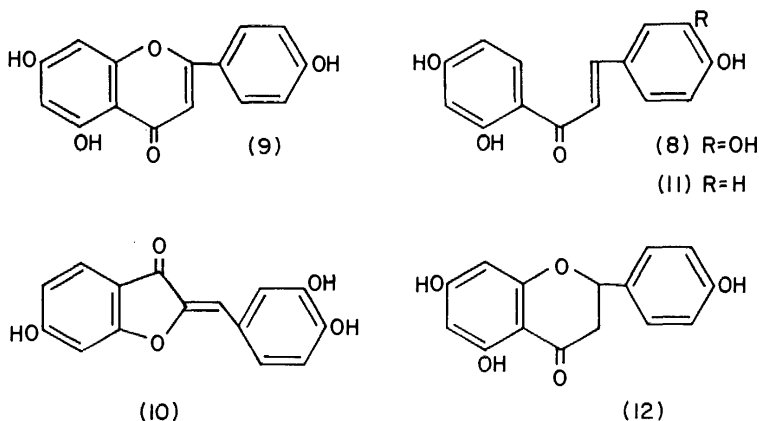






chalcone such as (2, R = OH) to a flavanone such as (12), the steps leading to plant products such as quercetin (4') and cyanidin (5, R,R' = OH) are oxidations which amount to introduction of oxygenated substituents or to dehydrogenation. The removal of a polyketide oxygen from ring-A was postulated to occur by reduction of an appropriate carbonyl, either in (1)<sup>9,10</sup> to form (7), cyclized with dehydration to (8), or possibly in the diketide precursor of (7) with or without dehydration before addition of the third C<sub>2</sub>-unit. As is usual in such cases the chemical theory outlines the process without defining exact intermediates. It is notable, however, that C<sub>6</sub>-C<sub>3</sub>-C<sub>4</sub> dihydro- $\alpha$ -pyrones are known plant products which may represent arrested stages in such a process. I note in passing the 2'-OMe isoflavone precursor (6) of rotenoids produced by migration of the B-ring.

Inheritance of flower colour provides a good illustration of genetic effects, and also provides a warning that the simple 'static' view of precursor-product relationships is often too simple a one.



Indication that the direction of transformation is flavonoid  $\rightarrow$  anthocyanin (without settling the details) is found in, for example, *Pharbitis nil* which has the genes Ca, C, A and R<sup>9</sup>. Recessive forms give neither pigment type; Ca gives flavonoids, Ca + C increases the amount, and Ca + C + R + A gives anthocyanin, more steps clearly being necessary for this type. Another notable point is that mutations of plants with change of oxygenation anthocyanin pigment in ring-B are almost always in the order: delphinidin (3 OH)  $\rightarrow$  cyanidin (2 OH)  $\rightarrow$  pelargonidin (1 OH). Since loss is more likely than gain of a stage, the oxidations must be in the reverse order, i.e. pelargonidin or a precursor is converted into cyanidin or a precursor and this into delphinidin. *Dahlia* is tetrasomic, so that each character for pigment production can occur up to four times presumably with increase of enzyme production with each genetic unit. The characters are: I, producing the flavone apigenin (9); Y, resulting in formation of butein (8) and probably also (10) and (11); A and B, which must be present for anthocyanin production, and which may differ merely in that B is more effective in enzyme production than A. The

anthocyanidin formed is either pelargonidin (5,  $R' = H, R = OH$ ) or cyanidin (5,  $R', R = OH$ ); its nature is determined in this plant not by a specific gene, but by the total of pigment genes (pelargonidin above a certain dosage of genes).

The gene Y can be explained on the lines above; it must control a reducing enzyme resulting in reduction of a polyketide carbonyl to carbinol at some stage (e.g. 7) before cyclization and aromatization. The analogous (11) could arise as in the anthocyanin story below; if so a prediction is that as the dosage of Y is increased the ratio of (8) to (11) should increase since reduction in the chain will be accelerated relative to hydroxylation of the ring. These chalcones are apparently not substrates for further conversions, apart from conversion to the aurone (10) which may be formed through the epoxide. The virtual disappearance of other pigments when Y is present indicates that the carbonyl reduction is a very efficient one.

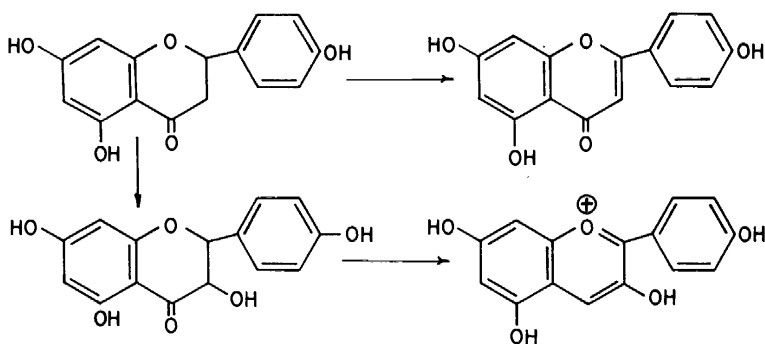
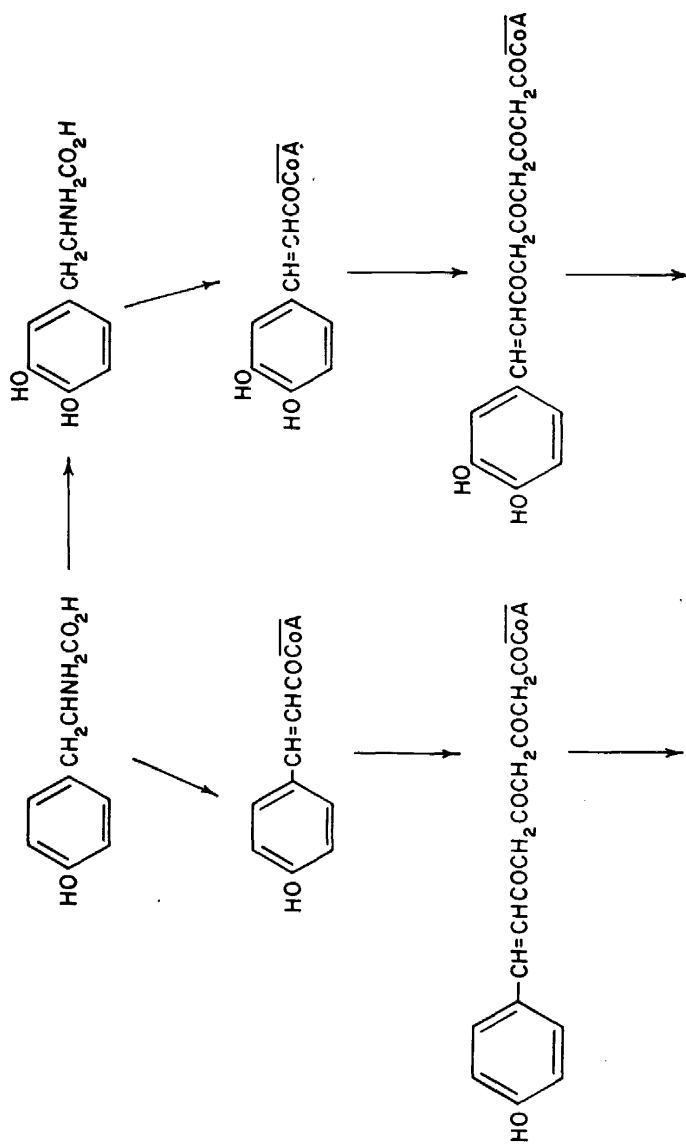


Figure 5

Lawrence and Scott-Moncrieff<sup>11</sup>, who did not have the advantage of the chemical model, arrived at the following genetic conclusions:

- (i) A<sub>1</sub> factors compete for a limited amount of starting material, the amount of anthocyanin increasing with dosage from A<sub>1</sub> to A<sub>4</sub>, and the amount of apigenin from I<sub>1</sub> to I<sub>2</sub>.
- (ii) B Y factors compete for a larger pool of starting material.
- (iii) A/B and Y factors compete among themselves for starting material, but there is no indication that the latter increases above a limiting value.
- (vi) B and Y show no increase in intensity above B<sub>1</sub> and Y<sub>1</sub>, but do increase in competitive power B<sub>2</sub>-B<sub>4</sub> and Y<sub>2</sub>-Y<sub>4</sub>.

The suggestion thus made that the pool of precursor (on our hypothesis probably *p*-hydroxycinnamoyl Coenzyme-A) differs according to the genetically conditioned withdrawal up to two different limits is puzzling. Presumably an intermediate pool can supply up to a certain maximum rate according to the rate of withdrawal in various directions and of maximum rate of synthesis, thus explaining the competition under (iii). The fact that B and Y show no increase above a dosage of one genetic unit indicates that the enzymes they control at this level are capable of dealing with all of the precursor available



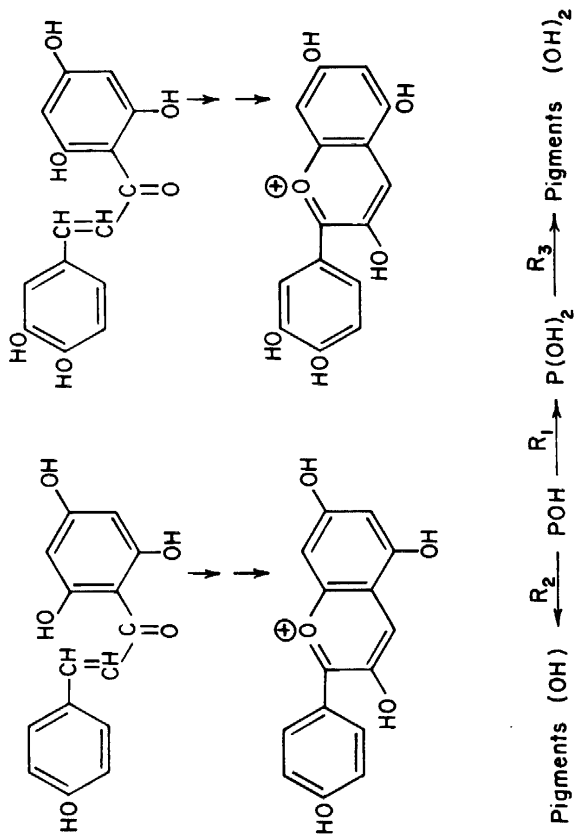


Figure 6.  $R_2 \gg R_1$  gives pelargonidin,  $R_3 \gg R_1$  gives cyanidin, normally  $R_3 > R_2$

to them. Higher dosages of gene, if they result in increase of the enzyme, will, however, increase the rate of conversion at the specific stage they control and hence will increase the competitive power for limited precursor. If the direct precursor of the main branch of pigments is the chalcone such as (2) and not the flavanone (12) with which it can be enzymically equilibrated, the enzyme controlled by I may dehydrogenate only (12) to the corresponding flavone apigenin (9), and (12) is effectively its specific precursor. Equilibration of (2, R = OH) and (12) may be slow compared with withdrawal of (12) and consequently increasing the latter rate by higher dosage of I produces no increase in apigenin, since all available precursor is being utilized anyway. The effect of A in increasing anthocyanin production up to  $A_4$  may be due simply to the fact that A is a poor producer of enzyme and even at full dosage may not be using all of the precursor available. Alternatively, it is conceivable that two routes to the flavanonol precursor of the anthocyanins could be involved, one (A) through the flavanone (12), and the other (B) directly from the chalcone; it is not yet clear that the same chemical classes of flavonoids are always produced in different plants by the same mechanisms.

Some practical results illustrating competitions are worth quoting (recessive genes are not shown):  $Y_1I_3$  gives little apigenin (due to I) and  $Y_2I_2$  gives none.  $A_1Y_1$  gives almost no anthocyanin (due to A) but this is detectable at  $A_4Y_1$ .  $B_1Y_1$  gives some anthocyanin (due to B) and  $B_1Y_2$  gives less. I (apigenin) competes strongly with A (anthocyanin) but complete suppression occurs only in extreme cases such as  $A_1I_{3-4}$ .  $BI_4$  produces less anthocyanin than  $B_1$ .

In some flowers, there are separate genes for anthocyanin types and in these cases the enzymes must be specific to ring-B hydroxylation patterns. From chemical models flavones might be alternatively derived by direct oxidations of chalcones, or by ring-closure to flavanones initially. Occam's razor is sometimes a rather blunt instrument in biochemistry.

In *Dahlia* the structure of the B-ring of the anthocyanin is determined in a non-specific way by the total demand of the genes for precursor. The principal occurrences are: cyanidin in  $A_4I_4$ ,  $B_1$ ,  $B_1I_2$ ,  $B_1A_2I_2$ ; pelargonidin (total or partial) in  $A_1Y_1$ ,  $B_1Y_1$ ,  $B_1A_1I_2$ ,  $B_1A_3I$ ,  $BI_3$ ,  $B_2$ . There is also an inhibitor gene H of Y and when this is present cyanidin is formed in proportion to the amount of inhibition; no inhibition gives pure pelargonidin. The results can be summed up by saying that the greater the total demand on precursors the greater the proportions of pelargonidin (5, R' = H, R = OH) to cyanidin (5, R,R' = OH). Since the direction of hydroxylation is mono-to di-, this oxidation must occur in some precursor at a rate which is presumably potentially independent of the presence of the pigment genes. A possible stage is the flavanonol since the only flavone formed by I is apigenin without luteolin (dihydroxy). The subsequent enzymes must, in this plant but not necessarily in others, be capable of transforming both the mono- and dihydroxy-series, but at different rates. The results can be explained if the rate with the dihydroxy-series is higher, resulting in cyanidin if the dihydroxy-precursor of this is present. If, however, with higher dosages of pigment genes, the withdrawal rate of precursor is faster than that of the hydroxylation reaction, precursors will be withdrawn at the mono-hydroxy stage, resulting in pelargonidin as shown in *Figure 6* ( $P = C_{15}$ -precursor).

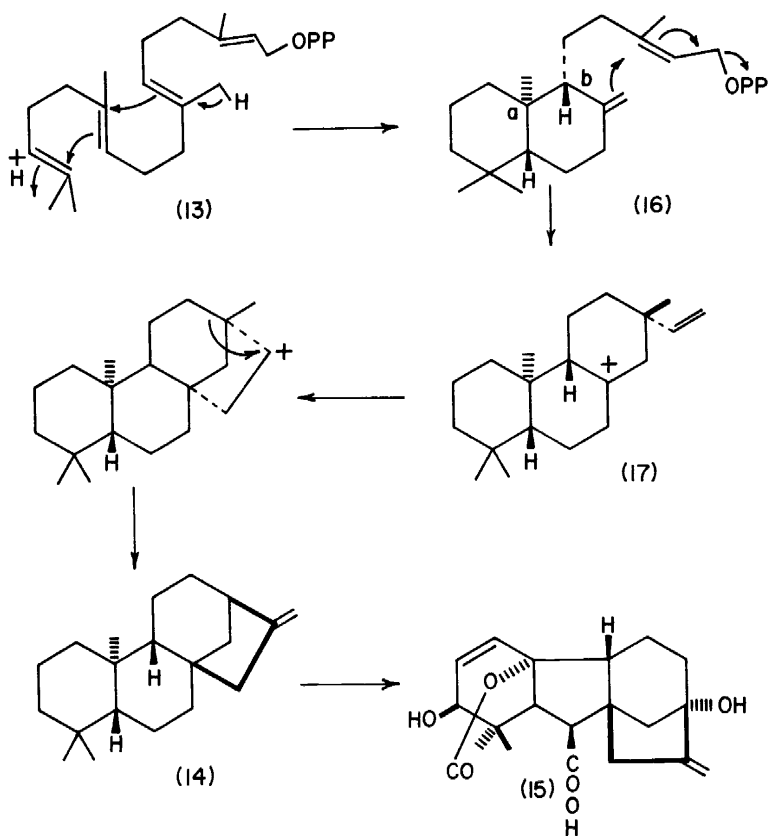
This example is quoted chiefly to demonstrate that a complex picture of related metabolites is not entirely influenced by simple precursor-product relationships, but by more subtle genetic influences which may be manifested in relative reaction rates.

Possible inferences from the results are that the more highly developed processes are: (1) cyanidin formation since this involves an extra hydroxylation stage; and (2) butein formation, since this involves an extra reduction stage. Apigenin formation is also an offshoot of the main process and may well have developed later than the main pigment route.

### DITERPENES

A different type of area, where changes are concerned with changes in carbon skeletons and stereochemistry is that of the diterpenes.

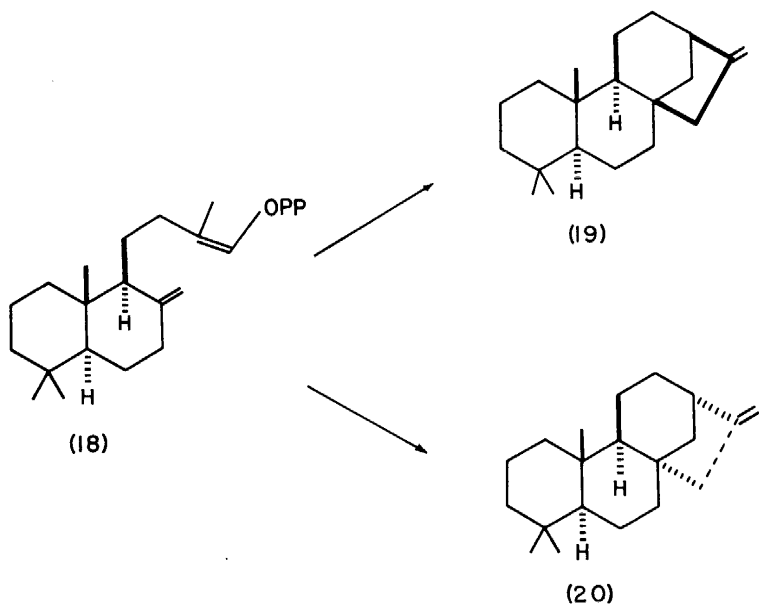
Farnesyl pyrophosphate is an important 'target molecule' since it is basic for diterpene, triterpene and steroid synthesis. Its reactions are so diverse however, that it is better to choose a much more restricted class such as the natural cyclic diterpenes derived from geranylgeranyl pyrophosphate which can virtually all be classified as offshoots of the biosynthetic process directed towards the 'target molecule' gibberellic acid (15). The biosynthesis



of (–)-kaurene (14) from geranylgeranyl pyrophosphate (13) is outlined below. We are not particularly concerned with the stages from (14) to (15) except to note that a series of oxidations at saturated centres are involved which, however ‘improbable’ on classical organic experience, appear to cause no difficulty.

In considering whether other diterpenes have diverged from this route (which seems very likely in view of the obvious structural relationships and the much more restricted types of structure found in diterpenes compared with the protean ring systems of the smaller  $C_{15}$ -sesquiterpenes) we must consider (i) absolute stereochemistry of precursors; (ii) occurrence of Wagner–Meerwein carbonium-ion rearrangements; (iii) specific oxidations.

Taking a comprehensive review as a non-selective basis<sup>12</sup> we have tabulated the known absolute configuration of bi- and tri- and tetracyclic diterpene derivatives. Considering at this stage only the centres (a) and (b) of (16) it can be said that most, but not all, bicyclic diterpenes do not have the (–)-kaurene configurations at (a) and (b), or are derived by migrations from precursors which do not have this configuration. They are mainly derived from the enantiomeric series (18). An even lower proportion of the tricyclic compounds [structurally related to (17)] have the (–)-kaurene configuration at these centres. The situation is completely different in the tetracyclic series; virtually all of the numerous compounds known—the alkaloids of the classes of garryfoline, atisine, lycotonine, the gibberellins, the grayanotoxins, cafestol and kahweol, enmeins, the beyerane group—are of the (–)-kaurene configuration and could have been derived from it or its immediate carbonium ion precursor. A very few are not so derived, including the interesting isomer of (–)-kaurene, (+)-phyllocladene (19) which occurs with (+)-kaurene (20) in *Podocarpus ferrugineus* and could be derived from the same bicyclic precursor in this tree:





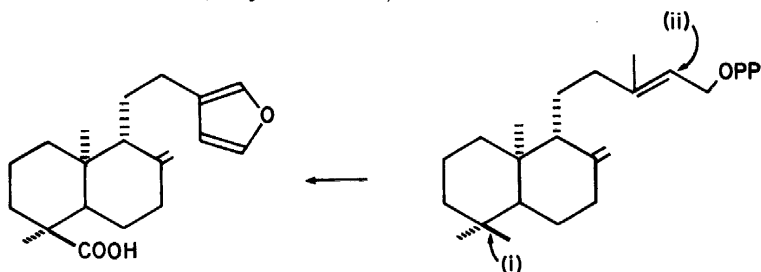
An explanation of the situation seems possible by considering what happens if mutations affect the absolute configurations of the ring-closures. The *relative* stereochemistry generated in the AB-ring-closure is that to be expected on chemical grounds so that presumably the enzymes required are quite 'probable' ones. The centres (a) and (b) have not been found in the less probable diastereoisomeric relationship. The diversion of almost all mono- and bicyclic diterpenes (but not quite all) from the gibberellic acid route is therefore hypothetically due to a reversal of the absolute configuration of the AB-ring closure and the adjacent centre. As a result the closure of the C-ring fails altogether in a number of cases since the gibberellic acid enzymes cannot act and various carbobicyclic products result, some of which contain other structures (oxide rings, carboxyl groups) which represent the development of fairly probable transformations of the arrested intermediates. When closure of the third ring does occur, either on the basis of modified or new enzymes, closure of the fourth ring does not occur, with the sole exceptions of phyllocladene (19) and (+)-kaurene (20), but alternative ways of disposing of the intermediate carbonium ion are developed, including migration, hydration and deprotonation reactions. Additional oxidation processes can occur. Each of these developments could be characteristic of a particular plant.

Assuming that production of (-)-kaurene is required for gibberellic acid biosynthesis, the means of producing it and its immediate carbonium ion precursor must be always available, unlike the other stereoisomers and consequently the other tetracyclic compounds are all derived from this ubiquitous precursor or its carbonium ion precursor.

To sum up: with reversal of stereochemistry at (a) and (b) if the products form at all they cannot be metabolized by the gibberellic acid route enzymes and alternative routes develop in the bicyclic and tricyclic series. Some genuine gibberellic acid intermediates also occasionally become converted into alternative products. Virtually all of the tetracyclic compounds could be derived by metabolism and rearrangement of (-)-kaurene or immediate congeners.

To note a few examples among many of divergences among which some may have phylogenetic significance, the reactions being shown in an arbitrary order.

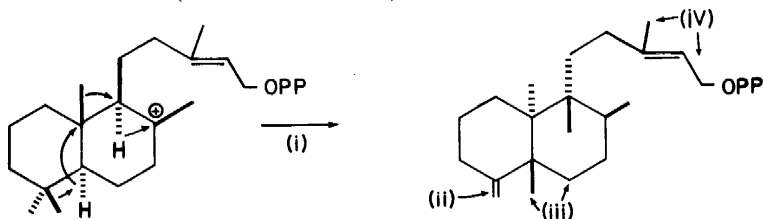
#### Gibberellic Acid Series (Polyalthic Acid)



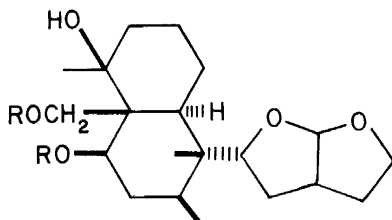
- (i) oxidation to  $\text{CO}_2\text{H}$ ,
- (ii) hydrolysis and oxidation to furan (via allylic oxidation?).

Both types of process are known in other cases, but represent the kind of extra stages which may be noted in assessing the significance of a constituent.

Enantiomeric Series (Clerodin Skeleton)

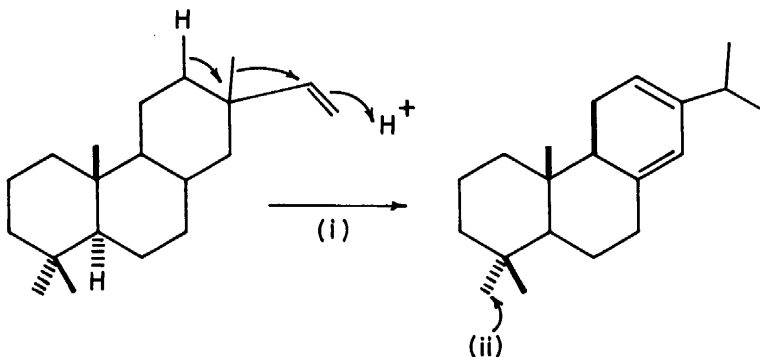


- (i) classical stereospecific carbonium ion rearrangement, involving four group-movements,
- (ii) epoxidation,
- (iii) acetoxylation (oxidation-esterification?),
- (iv) allylic oxidation by somewhat undefined routes to clerodin below.



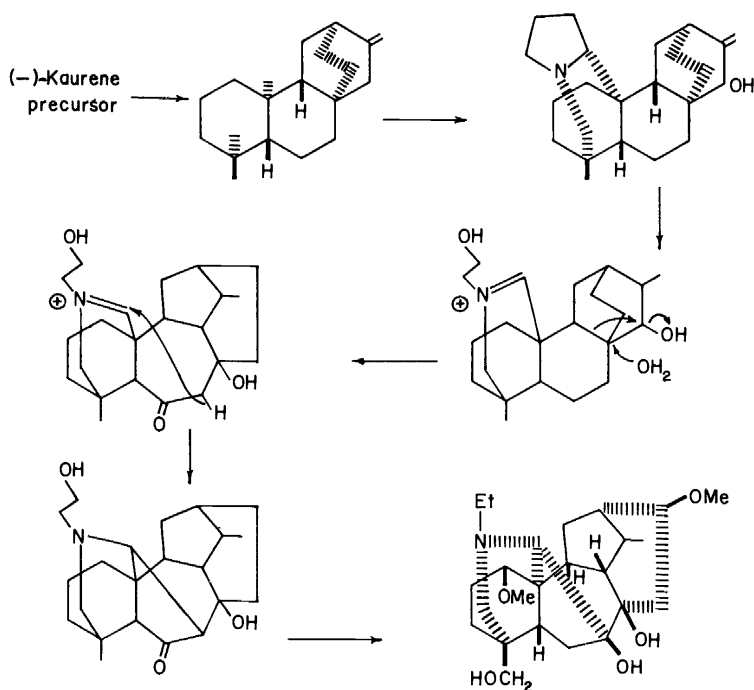
Many of these additional processes are 'improbable' and therefore likely to be rarely developed and highly characteristic.

Tricyclic Series (Abietanes) (Laevopimeric Acid)



- (i) acid catalysed rearrangement,
- (ii) specific oxidation.

## Tetracyclic Series (Atisine Lycoctonine)



The rearrangement is the most specific marker in this case; the choice of oxidized Me may also be significant.

This is a rather extreme case of skeletal rearrangement and oxidation and methylation, but despite the uncertainty about the order of events it is clear in that atisine is more primitive than lycoctonine by a definable number of stages. A number of rather 'improbable' oxidations and rearrangements are involved and may well be specific markers.

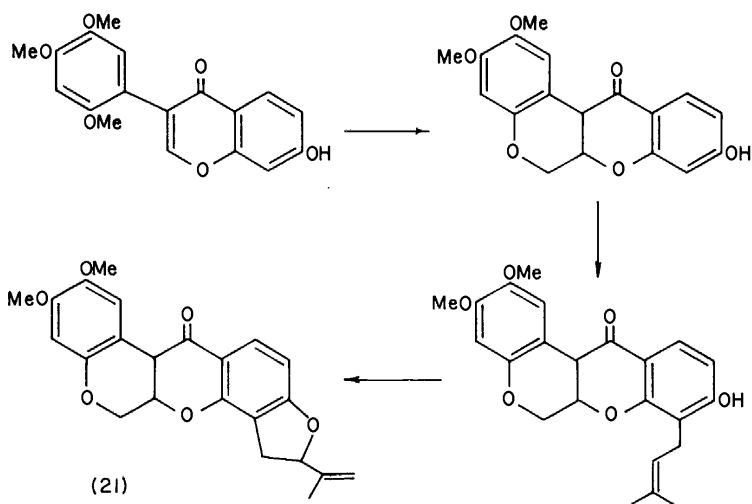
In general the structural points definable apart from stereochemistry concern the nature and number (both of which can be related to probability) of C—C and C—H bond migrations; the oxidation processes which can be fairly trivial (e.g.  $\text{CH}_2\text{OH} \rightarrow \text{CHO} \rightarrow \text{CO}_2\text{H}$ ) or more specific (e.g.  $\text{CH}_2 \rightarrow \text{CHOH}$ ) and the introduction of nitrogen (if it occurs) and other sidechains. In order to assess the full significance of such alterations in pathways a complete analysis of all diterpenoid compounds in a given plant (even minor ones) seems desirable.

Several other series of diterpenes such as the eremolactone series and the cembrene series, much more restricted in numbers, clearly have nothing to do with the gibberellic acid route and are therefore quite distinct markers.

A third type of biogenesis which should be particularly amenable to analysis involves compounds whose skeletons are elaborated from different biogenetic sources. Examples have already been discussed<sup>9</sup> in the variations of the  $(\text{C}_6\text{--C}_3)_x + \text{C}_{2y}$  routes. Additions or subtractions to the skeleton are

readily observed and can often be placed as divergences from a main route. An illustrative example is the series of rotenoids. They occur like the isoflavones on which they are based, chiefly in *Leguminosae* (Papilionatae).

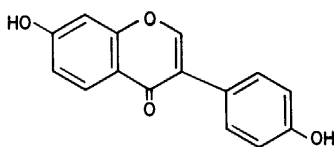
An outline of the later steps<sup>13</sup> in rotenone biosynthesis is:



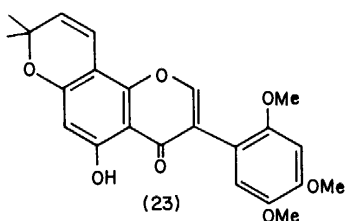
The biogenetic units in the initial main skeleton are the C<sub>6</sub>-C<sub>3</sub> (tyrosine) and C<sub>6</sub> (acetate-derived) units of the flavonoids. In the case of rotenone (21) removal of an oxygen from the phloroglucinol precursor presumably occurs as discussed for butein above. Migration of the aromatic B-ring is of the isoflavonoid type, and daidzin (22) occurs in rotenoids. This migration is much more frequent in compounds containing a 2'-oxygen, but whether it precedes or follows the migration is not clear. Methylation on oxygen, followed by oxidative ring-closure of the OMe, as already discussed would give the tetracyclic precursor above. Introduction of a C<sub>5</sub>-unit and oxidative ring-closure would then give rotenone (21). Other rotenoid constituents illustrate failures of stages or additional steps. The compound (23) represents failure of the OMe ring closure, and an alternative cyclization of the C<sub>5</sub>-unit. Homopterocarpin (24), apart from obvious deviations in the phenolic oxygen pattern, would represent failure of 2'-methylation, and reduction of the pyrone ring. The unsubstituted furan ring of (25) represents alkylation in a different situation and loss of the C<sub>3</sub>-portion of the C<sub>5</sub>-unit as already discussed. Another offshoot of the scheme is scandenin (26) with an extra terpene unit, no 2'-oxygen, no loss of oxygen from the phloroglucinol ring and oxidation of the pyrone ring.

It is possible to give some estimates of the probabilities of the steps involved in forming (21) from the mechanism point of view and from the frequencies in which they are encountered. Of the aromatic hydroxylations the least probable is the introduction of the 2'-OH, although this is also strongly linked with the isoflavone migration, itself significantly uncommon. Loss of the phloroglucinol OH, which requires as noted a specific reduction in an

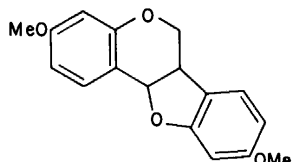
intermediate, is not uncommon, but does not occur in some congeners of rotenoids and is accordingly to be noted. Oxidations of the pyrone ring as in (26) are mechanistically acceptable, but are not common in natural products and accordingly significant. Undoubtedly the most significant step is the cyclization of OMe to form the pyran ring. The mechanism of this is not clear, and although predicted<sup>1,8</sup> on the basis of the natural formation of methylenedioxy from orthomethoxyhydroxy groups, does not have laboratory analogies. The least significant steps are the phenolic methylations, with some additional notability in the appearance of methylenedioxy groups.



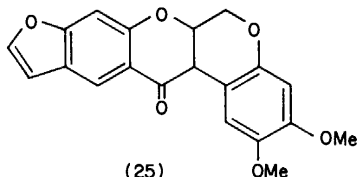
(22)



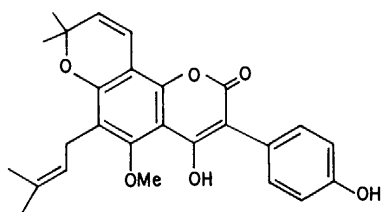
(23)



(24)



(25)



(26)

Introduction of a terpene unit into phenols, usually of the phloroglucinol or resorcinol type which are activated for reaction with carbonium ions, is not very common in flavonoids, but not unknown. It is clearly of significance, but the types of oxidative ring-closure (to dimethylpyran or isopropylidihydrofuran types), once the C<sub>5</sub>-unit is present, seem to be rather readily interchangeable alternatives. The presence of unsubstituted furan, as noted, required an extra step, but is not uncommon once the terpene unit is present. On the basis of such qualitative considerations it is possible in the case of a given constituent to say how 'probable' it is and therefore how significant as a marker.

As is obvious, the main problem with the approach is to decide whether differences between related series are due to gain of steps in one series, or loss in the other, i.e. to decide which way the evolutionary arrow is pointing. It seems unlikely that this difficulty will be resolved except by examinations of a number of biogenetically unrelated sequences to obtain a consensus of opinion.

A number of reviews of various aspects of the subject may be consulted<sup>1, 2, 3, 8, 9, 16</sup>.

I am greatly indebted to Dr J. Baldas for assistance in collecting information.

## REFERENCES

- <sup>1</sup> A. J. Birch, in *Chemical Plant Taxonomy*, Edited by T. Swain, p 141. Academic Press: New York (1963).
- <sup>2</sup> R. G. Huntsman in *Chemotaxonomy and Serotaxonomy*, Edited by J. G. Hawkes, p 21. Academic Press: London (1968).
- <sup>3</sup> J. D. Bu'lock, *The Biosynthesis of Natural Products*, McGraw-Hill: New York (1965).
- <sup>4</sup> A. Bracken, *The Chemistry of Microorganisms*, p 185. Pitman: London (1955).
- <sup>5</sup> A. J. Birch, 'Biosynthetic intermediates in polyketide biosynthesis', from *Biosynthesis of Aromatic Compounds*, p 3. Pergamon: Oxford (1966).
- <sup>6</sup> e.g. *Biogenesis of Natural Compounds*, Edited by P. Bernfeld. Pergamon: Oxford (1968).
- <sup>7</sup> A. J. Birch, R. E. English, R. A. Massy-Westropp, M. Slaytor and H. Smith, *Chem. & Ind.* 204 (1947); *J. Chem. Soc.* 365 (1958).
- <sup>8</sup> A. J. Birch, *Biogenesis delle Sostanze Naturali* (Accad. Naz. Lincei, Roma), 79 (1964);  
A. J. Birch and M. Slaytor, *Chem. & Ind.* 1524 (1956).
- <sup>9</sup> A. J. Birch, 'The biosynthesis of flavonoids and anthocyanins', Plenary Lectures, p 73, 17th Internat. Congr. Pure & Appl. Chem. (1959). Butterworths: London (1960).
- <sup>10</sup> A. J. Birch and F. W. Donovan, *Austral. J. Chem.* 6, 360 (1953).
- <sup>11</sup> W. C. J. Lawrence and R. Scott-Moncrieff, *J. Genetics*, 30, 155 (1935);  
cf. W. C. J. Lawrence and J. R. Price, *Biol. Rev. Cambridge Phil. Soc.* 15, 35 (1940).
- <sup>12</sup> R. McCrindle and K. H. Overton, *Advances in Organic Chemical Methods & Results*, Vol. 5, p 47. Interscience: New York (1965).
- <sup>13</sup> L. Crombie, P. M. Dewick and D. A. Whiting, *Chem. Commun.* 1182 (1971).
- <sup>14</sup> A. J. Birch and H. Smith, *Chem. Soc. Spec. Publ.* 12, 4 (1958);  
A. J. Birch, Myint Maung and A. Pelter, *Austral. J. Chem.* 22, 1923 (1969).
- <sup>15</sup> H. G. Floss and V. Mothes, *Phytochemistry*, 5, 169 (1966).
- <sup>16</sup> J. B. Pridham and T. Swain, *Biosynthetic Pathways in Higher Plants*, Academic Press: New York (1965);  
J. B. Harborne, *Comparative Biochemistry of the Flavonoids*, Academic Press: New York (1967);  
M. Calvin, *Chemical Evolution*, Clarendon: Oxford (1969);  
*Progress in Phytochemistry*, Edited by L. Reinhold and Y. Liwschitz, Vol. 1. Interscience: New York (1968).