Journal of the National Science Council of Sri Lanka, 1 (1972), 19—29

The Construction of Biosynthetic Hypotheses

A. J. BIRCH,

Research School of Chemistry, Australian National University, Canberra.

Relations between structure and biosynthesis have stimulated speculation since early days of natural product work. The subject was, and still is, an art rather than a science, one of its most appealing aspects being that it involves creative speculation, the results of which can ultimately be tested. It involves induction, rather than deduction of the type which marked the early work by classical biochemists. It is characteristic of the organic chemist rather than the biochemist. The great advantage of such speculation is that, if successful, it can suggest appropriate biochemical experiments in connection with complex structures which might be very laborious to tackle by trial and error. The big disadvantages are that with the exception of the acetate polyketide theory discussed below, the units are usually large, and theory often does not indicate how they arise, and that the exact order in which a series of transformations occurs often is not suggested.

Success in such speculative exercises cannot be based on automatic rules : it is dependent on an ability to recognise the biogenetic units in molecular skeletons, despite obscuring subsequent processes. This in turn depends largely on the ability to postulate appropriate chemical reactions, which may well not be known biochemical ones at the time, for the junctions of the units and for subsequent changes. A very good knowledge of organic mechanisms, and a feeling for the compatibility of the postulated processes with what is known of biochemical mechanisms are requisites. The classical biochemist frequently knows too little about the former, and is perhaps unduly inhibited in going far afield in the latter.

A survey of some previous approaches reveals more of the requirements for success than would an attempt at a systematic treatment of the subject.

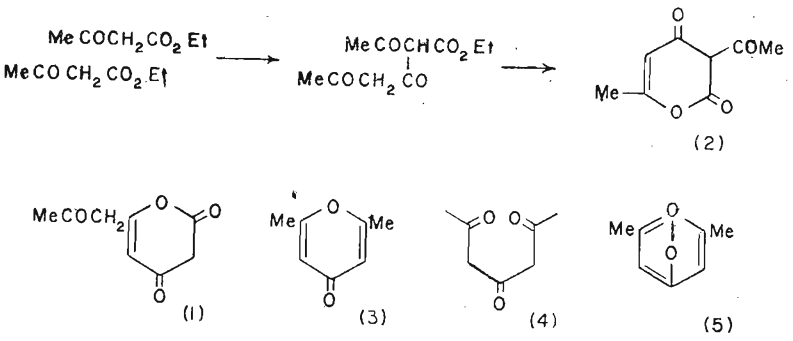
The first major attempt at a biosynthetic correlation, the polyketide hypothesis[1], was developed in useful detail only some 45 years after its initial postulation, which was itself completely sterile. Its initial failure is a good example of

how the proverbial baby can be thrown out with the bath water. The essence of truth was obscured by unnecessary accretions due to attempts to extend it beyond its scope to an all-embracing philosophy, and to be original at all costs no matter whether or not the point at issue was genuinely relevant to the hypothesis. The fact that it could be used, (although with the exception of orsellinic acid was not explicitly used) to explain a number of natural structures very well was neglected, apparently because of the obvious straining to explain rather fancifully and unconvincingly other structures claimed to be within its scope. Its exposition was vague and in some places contradictory. Two lessons emerge : the hypothesis must be very clearly stated and its limitations circumscribed by the theoretical support available. Later work[2] in the area is an even more explicit example of the approach not to take : to start with a fixed idea that, for example, carbohydrates of any size or chain shape are naturally available, despite lack of evidence that this is so, and then to postulate unlimited enzymic specificity in oxidations, reductions, or ring-closures. It is then possible to make anything on paper, and the results are meaningless in terms of suggesting fruitful biochemical experiments.

The original author of the polyketide hypothesis, J. N. Collie, in fact began[1] with a very good and sound idea, based on his laboratory experiences with " dehydracetic acid " and some of its derivatives, namely that β-polyketones can ring-close or react with each other to give phenolic compounds by aldol or C-acylation procedures (from esters) and that in this way a number of natural products could be structurally explained. If his paper of 1907[1] had been followed up logically, as was the alkaloid biosynthesis paper of Robinson in 1917[3], it could have marked a major step in biosynthetic understanding.

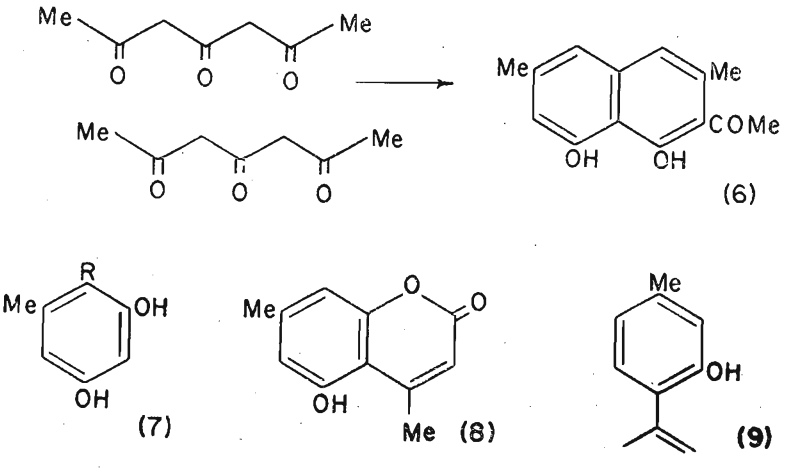
The only subsequent publication was in a textbook[4] by A. W. Stewart, a personal friend of Collie, and the account, from general acknowledgements to Collie and with many paragraphs marked as direct contributions from him, must be considered as authoritative of his views. The clear stream of the original idea is lost in this account in a maze of muddy ancillary hypotheses.

The starting-point of the initial hypothesis was the experimental investigation of dehydracetic acid then wrongly formulated as (1) [actually it is (2) from a branched rather than a straight-chain, and the error is of little consequence to the theory]. The substance is derived from two molecules of acetoacetic ester, and hence from four molecules of acetic ester.



The action of acid converts it into dimethylpyrone (3) presumably via the " polyketide " (4). The action of alkali on (4) led first to a monocyclic and then to the bicyclic phenol (6) as shown. Under different conditions dehydracetic acid on reaction with alkali gave orcinol (7, R=H) and orsellinic acid (7, R=CO_2H) known from natural sources. The resemblance of (6) to natural phenols, and the identity of (7) with a natural substance led Collie to the idea that such " polyketens " or " polyketides " might be intermediates in phenol biosynthesis. Unfortunately he did not discuss other natural examples.

A further experimental observation was the condensation of orcinol with acetoacetic ester to (8), and decarboxylation of this to (9), which has, as he points out, the skeleton of the monoterpene thymol. He did not in any explicit terms attempt to extend the idea to terpene biosynthesis in general. If it can be assumed that he was thinking in terms of polyketide origins in acetate then this is the first suggestion of such an origin for a terpene.



It is not however clear from his or Stewart's summary whether or not he had any clear idea of the involvement of acetate units, although this might have been assumed from the experimental origins of the work. In the later discussion in Stewart's book the question is largely ignored, apart form a statement[4] (p 276) " keten can polymerise to long chains which then add water to form polyketides " The discussion is taken up almost entirely, however, with the possible origins of polyketides from carbohydrates by directed dehydration. There is also the suggestion that this may be reversible by reverse hydration of polyketide enols, incidentally an unlikely process mechanistically and without laboratory analogy as Collie admits : " it must be frankly confessed that up to the present our laboratory methods have failed to bring about either of these conversions " (i. e. formation of polyketide from carbohydrate or the reverse.)

The fact that very unusual carbohydrates would be required, and were not known, was also ignored. Whether the keten units were thought to arise from acetate, and whether the keten idea was tacitly dropped in favour of carbohydrate origins from formaldehyde, of which there is discussion, is not clear. Certainly the confusion of ideas could not have led to much faith by readers in the validity of any of them.

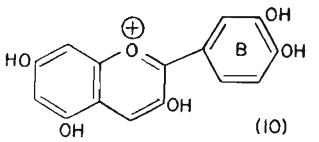
Another confusing aspect is the unnecessary attachment of rather bizarre hypotheses which really have nothing to do with the main one, such as a new formulation of carbohydrates, literally as hydrates of a carbon chain " these water molecules envelop the carbon atoms and that the three atoms of the water molecule rotate round the carbon atom to which they are related ".

$$\begin{array}{cccc} C\text{——} & C\text{——} & C\text{——} & C \\ (H_2O) & (H_2O) & (H_2O) & (H_2O) \end{array}$$

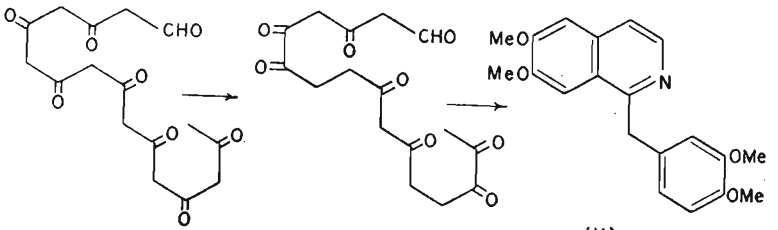
This was combined with a rejection of the " antiquated conception of directed valency ". If the polyketide idea was classified by readers with such ideas, it is no wonder it was disregarded.

An essence of the polyketide theory is that it places oxygens and ring-closures in 1, 3-positions to each other. Examples of straining the theory to fit inappropriately substituted examples are given by Stewart[4]., presumably with Collie's acquiescence. The anthocyanins and flavonoids are derived from an unknown C_{15}-sugar " the reaction may be traced directly back to a carbohydrate chain without requiring the intermediate formation of a polyketide derivative at all". This statement is presumably due to the extremely frequent occurrences of 4-hydroxy or 3, 4-dihydroxy B-rings, which do not fit the hypothesis, and the complete lack of 3-hydroxy and 3, 5-dihydroxy B-rings which do (compare cyanidin 10). Robinson later correctly used this oxygenation pattern to relate

this part of the molecule to many known C_6-C_3 compounds. Similarly, the alkaloid papaverine (11) was postulated by Collie, in a forced manner, to arise as follows, the α-dicarbonyl chain being formed arbitrarily " by the usual process of dehydration and rehydration ".



(10)

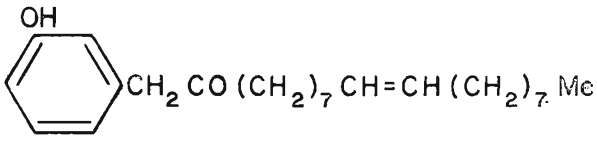


(11)

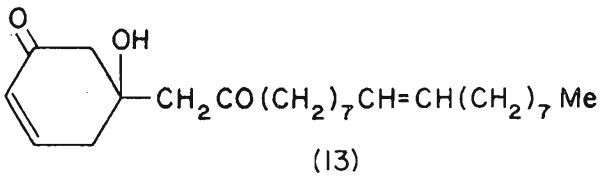
Our own approach to the polyketides taken in 1951-52[5], which owed nothing to Collie, since we were unaware of his work, or to Robinson who first published[6] his ideas after our own (1955), was based firmly on mechanistic considerations. It was fortunate in the availability of two major pieces of information from biochemical sources. One was the origins of some natural benzenoid compounds (particularly 4-oxygenated compounds such as tyrosine) from shikimic acid. The other was a knowledge of the intervention of " acetate units " (acetyl coenzyme-A) in the biosynthesis of fats and steroids.

Our first recognition of polyketide origin was with campnospermonol (12), the side-chain of which, up to the carbonyl, is clearly oleic acid (although the double bond stereochemistry was unknown). Since this acid arises from acetate units, the question was whether the rest of the molecule could come from a continuation of the process, invoking only mechanistically probable processes. Pencil and paper showed that it could arise as shown below, invoking reactions (a) aldol cyclisation (b) reduction of a carbonyl (c) dehydration of β-hydroxyketones and (d) decarboxylation of a β-keto-acid. The intermediate (13) was later isolated as a natural precursor of campnospermonol ;

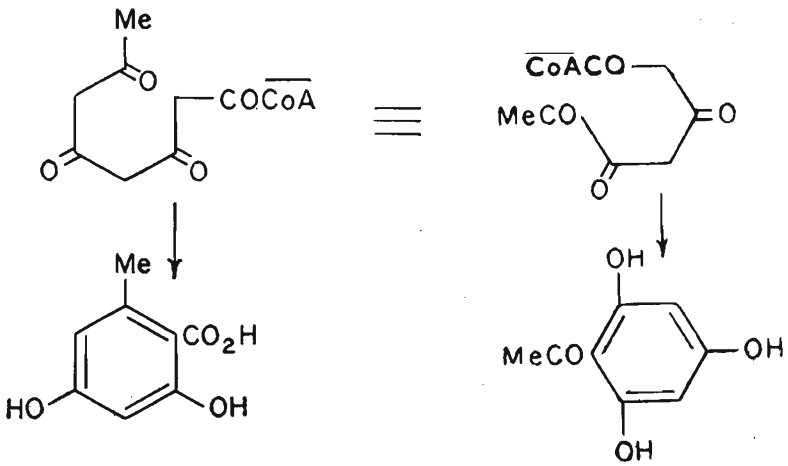
the theory of course did not indicate the order of some of the steps. Proceeding from this to simpler cases, orsellinic acid and acetylphloroglucinol were noted as the prototypes of a series of homologues and analogues, and could be formulated from the same precursor as follows :



(12)

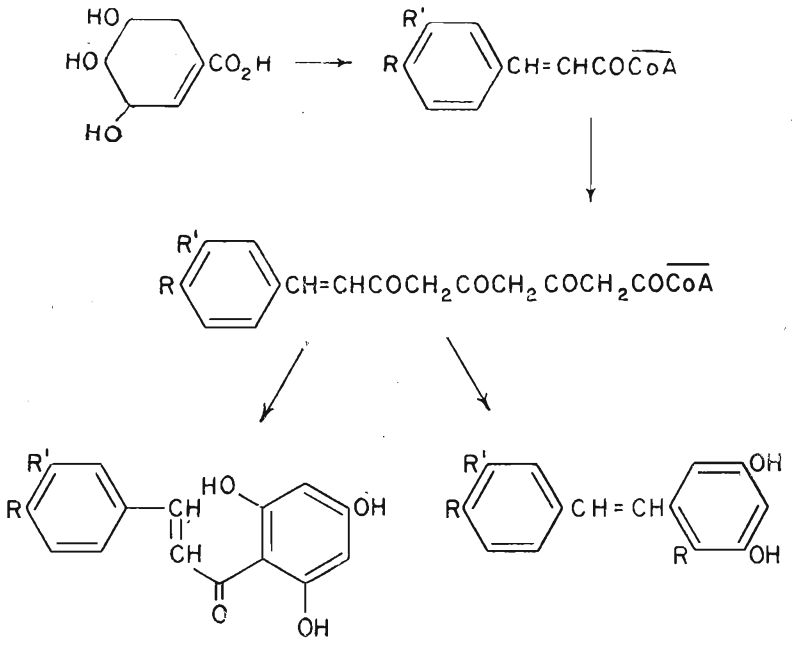


(13)



Loss of expected oxygen through specific reduction of a carbonyl in an intermediate, and the addition of other oxygens by o,p-oxidation readily accomodated other compounds. The hypothesis was rapidly tested, using radioactive tracers in moulds and plants, and shown to be essentially correct. Some 3000 natural structures can now be explained as wholly or partially polyketide. The details are too well known to merit repetition.

It is worth noting, however, that almost the first use of our hypothesis was to suggest for the first time the correct origin of the flavonoids and anthocyanins from a C_6-C_3 (shikimate) unit and three " acetate " units, the identical intermediate being convertible into the plant stilbenes :

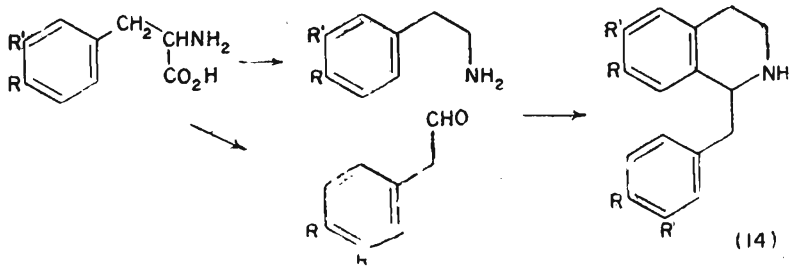


A completely logical approach, using only valid chemical reactivities, thus led directly to what was later shown to be the correct explanation of a long-standing puzzle and as a bonus provided the answer to another.

The still somewhat arbitary assumptions that had in general to be made, and they seem reasonable, is that enzymes determine the number of units involved, the exact directions of cyclisation processes, the situations of any carbonyl groups reduced, and of oxygen introductions. No type of reactivity invoked was novel from the chemical point of view.

Robinson's classical paper of 1917[3] on the biosynthesis of alkaloids is a fine example of a pregnant speculation, despite the fact that many of the details were later shown to be incorrect. It was valid and fruitful because of the main point made with numerous detailed examples : that a logical application of only two reactions, the generalised aldol condensation and the generalised reaction of amines with carbonyl compounds, could lead to a convincing

picture of the relations of many alkaloid structures to one another and to those of possible precursors. The predicted and proven origin of benzylisoquinoline alkaloids (e.g. 14) is shown. It probably succeeded in part because of the rather large size and recognisable character of many of the units involved (e. g. C_6C_2N, C_6CCO etc.) and the fact that they are marked out in the relatively permanent main skeletons by C and N, and are not so dependent on oxygen positions on a skeleton, as with the polyketides, a pattern readily obscured by oxidation-reduction processes. It was convincing however largely because its premises were clearly stated and extensively exemplified.

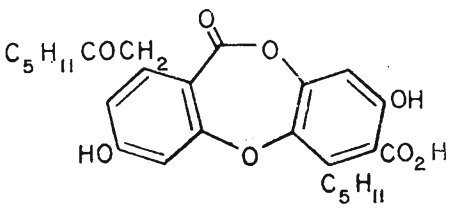


Finally, as a further example of the pitfalls of hypothesis, let us examine the C-methylation and propionate theories.

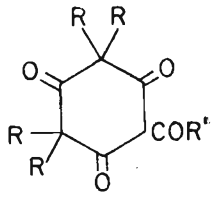
Many lichen products, such as (15) indicate from their structures an acetate-polyketide origin, despite some changes of an obvious nature, in this case an oxidative ring-closure. The number of such compounds was so great, that the basic theory seemed obviously correct. There were however some " extra " substituents inexplicable on this basis. By analysing all of the structures in a review the following distribution of substituents on the units (16) was found[7]. The side-chain always has an odd number of carbons, oxygens are always found in the 2, 4-(orcinol) positions and there are several " extra " oxygens at the 4- or 5- positions (and one Cl) which could be introduced by mechanistically acceptable processes. The problem was that 17 of the 57 structures contained a C_1-unit (as CH_3 CHO or CO_2H) at the 3-position. I recall that in a lecture in Liverpool in early 1952, Dr. F. M. Dean raised this as a valid objection to the theory. In conjunction with other evidence, such as the C-methylated acylphloroglucinols (17) (R=H or Me), the structures suggested the introduction to carbon of a C_1-unit, and although it was then not a known biochemical process, the question arose as to whether this could occur through trans-methylation

with methionine, known to occur with O-, S- and N-methylation. This expectation arose from the known mechanistic resemblance of such hetero-atom reactions to alkylations of phenol-enol systems, and laboratory analogies of formation of (17, R=Me) from acylphloroglucinols.

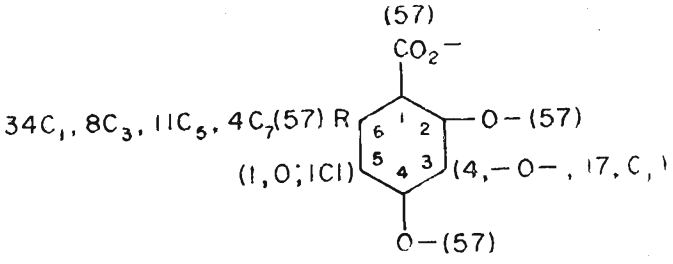
The hypothesis was rapidly supported[8] by using [^{14}C]-Me methionine with the mould metabolite mycophenolic acid, which contains both OMe and a similar C-Me, which are found to be equally labelled. Many other examples followed.



(15)



(17)



(16)

Position : 1. $26CO_2H$; $31CO_2-$:

2. 55OH ; 20Me :

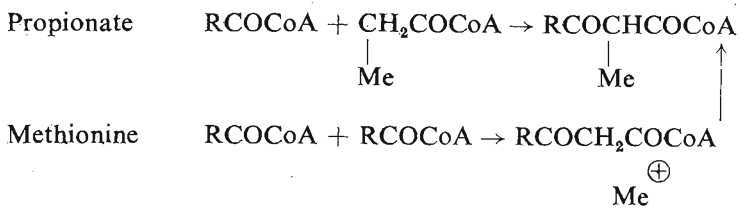
3. 10H ; 3-OCO- ; 9Me ; 6CHO ; $2CO_2H$:

4. 11OH ; 20OMe ; 26-OCO-:

5. 1Cl ; 10H :

6. 33Me ; $1CH_2OH$; $8C_3H_7$; $10C_5H_{11}$; $1C_7H_{15}$; $3CH_2COC_5H_{11}$ $1CH_2COC_3H_7$.

In 1954, at a Gordon Conference, I discussed the acetate-polyketide hypothesis, and R. B. Woodward who was present, and was then working on the structure of magnamycin, suggested the additional intervention of "propionate" to explain the C-methylated units in macrolides, as did Robinson in 1955 for mycolipenic acid. The final structural consequences of the two routes would be the same, as shown by the prototypes :



The later discovery of the intervention of malonate derivatives is irrelevant so far as the principle is concerned.

In fact both routes were found to occur in different cases, and " Occams razor " is rather dangerous as a tool in biosynthetic speculation ; the existence of one valid explanation does not rule out another possible one under different circumstances.

A second type of methionine C-methylation has also been observed. The classical one involves transmethylation and for example $^{14}CD_3$ is transferred from $^{14}CD_3$-methionine[9]. The observed methylation of an unactivated C=C as in the side-chain of ergosterol, would not have been predicted by the original approach, since the anionoid reactivity would not have been considered sufficient, as it is with C=C-OH. The mechanism was found to be different, transfer occurring of $^{14}CD_2H$, with loss of D. The intermediate is undoubtedly a cyclopropane, and the reaction resembles a carbene addition rather than a carbonium ion alkylation[10]. The initial ideas at least suggested the correct biochemical experiments, leading even to the discovery of an unpredicted process. In this area, any hypothesis is better than none, provided it leads to further experimental investigations.

Investigations of natural product structures are still of great interest in leading to new biosynthetic processes, or variants or refinements of old ones.

References

1. Collie, J.N., (1907) *J. Chem. Soc.*, **91**, 787, 1806; *Proc. Chem. Soc.*, 1907, 230.
2. Hall, J.A., (1937) *Chem, Reviews*, **20**, 305.
3. Robinson, R., *J. Chem. Soc.*, (1917) **111**, 762, 876.
4. Stewart, A. W., (1920) *Recent Advances in Organic Chemistry*, Longmans-Green, 4th Ed.
5. Birch, A.J. and Donovan, F.W., *Aust. J. Chem.*, (1953) *ā*, 360.
6. Robinson, R., *The Structural Relations of Natural Products*, Oxford, (1955).
7. Birch, A.J., Massy-Westropp, R.A. and Moye, C.J., *Aust. J. Chem.*, (1955) **8**, 539.
8. Birch, A.J., English, R.E.. Massy-Westropp, R.A., Slaytor M. and Smith, H., *Chem. Ind.*, (1957) 204 ; *J. Chem. Soc.*, 1958, 365.
9. Alexander, G.J., Gold A.M. and Schwenk, E., (1957) *J. Amer. Chem. Soc.*, **79**, 2967, 4554.
10. Lederer, E., (1964) *Biochem. J.*, **93**, 449.