

INTERNATIONAL UNION OF  
PURE AND APPLIED CHEMISTRY  
DIVISION OF ORGANIC CHEMISTRY  
in conjunction with  
THE SWEDISH NATIONAL COMMITTEE  
FOR CHEMISTRY

# THE CHEMISTRY OF NATURAL PRODUCTS

4

*Main Lectures presented at the*  
**FOURTH INTERNATIONAL SYMPOSIUM ON**  
**THE CHEMISTRY OF NATURAL PRODUCTS**  
*held in Stockholm, Sweden*  
*26 June—2 July, 1966*

LONDON  
BUTTERWORTHS  
1967

ENGLAND: BUTTERWORTH & CO. (PUBLISHERS) LTD.  
LONDON: 88 Kingsway, W.C.2

AUSTRALIA: BUTTERWORTH & CO. (AUSTRALIA) LTD.  
SYDNEY: 20 Loftus Street  
MELBOURNE: 473 Bourke Street  
BRISBANE: 240 Queen Street

CANADA: BUTTERWORTH & CO. (CANADA) LTD.  
TORONTO: 1367 Danforth Avenue, 6

NEW ZEALAND: BUTTERWORTH & CO. (NEW ZEALAND) LTD.  
WELLINGTON: 49/51 Ballance Street  
AUCKLAND: 35 High Street

SOUTH AFRICA: BUTTERWORTH & CO. (SOUTH AFRICA) LTD.  
DURBAN: 33/35 Beach Grove

*The contents of this book appear in*

**Pure and Applied Chemistry, Vol. 14. No. 1. (1967)**

*Suggested U.D.C. number: 547.9(063)*

©

International Union of Pure and Applied Chemistry  
1967

# **THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE**

**KARL FOLKERS**

*Stanford Research Institute, Menlo Park, California, U.S.A.*

## **INTRODUCTION**

When invited to present this lecture I was assured that this subject would provide a unique opportunity for perspectives. Not only was this correct, but this subject has been a unique trial for decisions on which products to exclude and include; I have excluded antibiotics, but pay tribute to their vast importance in medicine.

As a policy, contributions directly or indirectly from the chemistry of natural products are interpreted broadly, and I decided to minimize historical aspects and to maximize recent and current aspects of chemistry and medicine. Because most of us are interested in the future, I have included trends and medical objectives of great significance for new challenge.

I should like to give first priority to detailing the contribution of Swedish chemists to recognition of the impact of natural product chemistry on medicine.

## **CONTRIBUTIONS FROM SWEDEN**

It is evident that the pharmaceutical industry bridges the basic chemistry and the medical use of drugs and makes its own vital scientific contributions. The Swedish pharmaceutical industry also fulfils this commendable role for world health. There would be few, if any, new drugs in medicine today without this industry.

Research by Jorpes of the Karolinska Institute advanced Heparin, a mucopolysaccharide, and Vitrum introduced it for medical use. Up to 25 000 patients are treated with Heparin for thromboembolic diseases in Sweden each year. After its introduction, not only in Sweden but in other countries, the mortality of patients suffering from thromboembolic diseases has dropped from 15 per cent to 0·5–1·0 per cent. Heparin has also greatly lessened postthrombotic sequela and has benefited patients in operations with the heart-lung machine.

The researches of von Euler, Nilsson-Ehle, and Erdtman on a compound, later named gramine, and found in x-ray mutants of barley, led to the synthesis of 2-(dimethylaminomethyl)-indole, which differed from gramine by being isomeric and, surprisingly having local anaesthetic activity. Löfgren and Goldberg, aided by the serendipity, continued the research, and Astra ultimately marketed Xylocaine. It is estimated that over one-half million injections of Xylocaine are given daily in approximately eighty countries of

## KARL FOLKERS

the world. Xylocaine has been considered the world's leading local anaesthetic.

Lehmann of Sahlgren's Hospital in Gothenburg recognized that salicylic acid had an interesting effect on the metabolism of the tubercle bacillus. Then, he studied a larger number of related compounds for inhibition of the growth of the tubercle bacillus. *p*-Aminosalicylic acid, known as PAS, had promising inhibitory effects on experimental tuberculosis. With the contribution from Ferrosan, PAS became, and still is, one of the important drugs in the treatment of tuberculosis. It is estimated that the world production of PAS is about three million kilograms per year. Historically, the first successful test with PAS in clinical tuberculosis was made before the first clinical trials with streptomycin.

Dextran, which consists of polysaccharides, stemmed from the research of Grönwall and Ingelman of Uppsala and was introduced by Pharmacia into medicine. These polysaccharides were the first acceptable plasma substitutes, and their medical significance, especially in disasters and wars, is obvious. The number of patients treated with some form of Dextran approximates one million a year and usage is increasing. Many such patients could not have been saved at all before the availability of Dextran.

Kabi has succeeded in producing Streptokinase of high purity which is another step forward in the medical treatment of thrombosis. In appraising the medical potential of Streptokinase, one may estimate that a hospital of about 1000 beds might have a total of about 300 thromboembolic cases per year, including acute arterial occlusions, deep venous thrombosis, and myocardial infarctions, all of which could be eligible for Streptokinase treatment.

Biochemistry in Sweden is world renowned and will contribute to future medical therapy, and with the participation of the Swedish pharmaceutical industry. One can observe new possibilities, including the prostaglandins of Bergstrom and his colleagues.

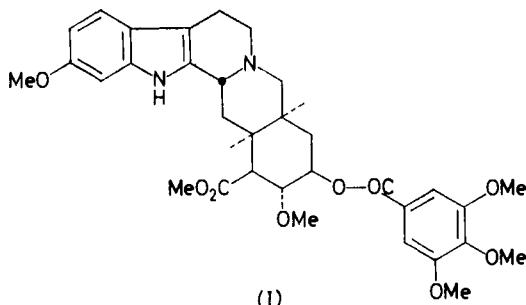
## TRADITIONAL ALKALOIDS

Of all the natural products studied by chemists in the last century, perhaps alkaloids dominate this chemistry, particularly if one takes into consideration the companion medical aspects. Although alkaloids have made a valuable contribution to medicine over many decades, one may observe that the contribution of alkaloids to medicine could diminish in the future, and that the landmark contributions to medicine will increasingly stem from investigations of basic biochemistry. This observation about alkaloids is based upon the increasing attention to rigorous requirements for safety in governmental regulations on drugs. Also, the alkaloid chemist faces formidable competition from the synthetic chemist and biomedical investigator who have biochemical theory and greater flexibility of structural variation at their command. Nevertheless, the contributions of alkaloids have not ended and may be expected to continue. The story of the *Rauwolfia* alkaloids has been told, and I only wish to pay tribute to their contribution to medicine. The use of *Rauwolfia* extracts extended over centuries, and credit is due to chemists and pharmacologists in India for their early research on the

## THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

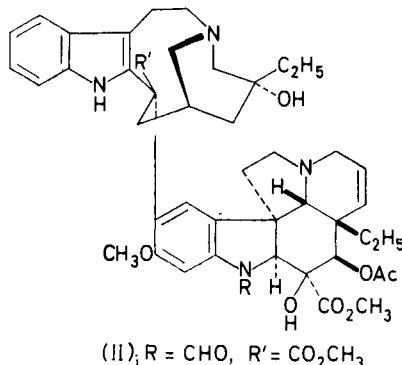
alkaloids of *Rauwolfia serpentina*. The use of *Rauwolfia* alkaloids began in Western medicine as recently as 1953.

Reserpine (I)<sup>1</sup> became widely used in medicine in mental conditions, including tension and anxiety, as well as in the treatment of some forms of hypertension. The introduction of chlorpromazine took place about a year later. The importance of *Rauwolfia* alkaloids in today's medicine may be



judged by sales of pharmaceutical preparations at 55 million dollars in 1963. Psychopharmacological agents have expanded in sales to over 200 million dollars; 18 million people afflicted with certain mental conditions are treated each year in the United States.

Also, in the 1950s, the periwinkle plant, *Vinca rosea* Linn., was studied because of an interest in diabetes. However, extracts of this plant caused leukopenia rather than hypoglycemia in animals. Eli Lilly and Company found such extracts to be active against experimental leukemia in mice. The alkaloid which stemmed from these observations is now known as vinblastine (II)<sup>2</sup>. Since 1960, a few hundred publications have appeared on medical studies with vinblastine, particularly in the field of cancer. While no



known drug, whether natural or synthetic, is as effective in the treatment of cancer as desired, each new oncolytic drug, which has some positive effect in the treatment of clinical neoplasms, is another step of encouraging progress.

## KARL FOLKERS

It appears reasonable today to consider treatment with vinblastine sulphate (II) in those cases of cancer, which include the following conditions: (1) lymphomas, Hodgkin's disease, lymphosarcoma and reticulum cell carcinoma; (2) monocytic leukemia; and (3) carcinomas of the breast, which have not responded to other treatments.

Lower dosage and longer treatment for the more slowly responding tumours, such as carcinomas of the breast and bronchus, have been advised. Vinblastine seems to lack cross-resistance with other therapies. Leukopenia is the dose-limiting factor. Although some oncolytic drugs, including vinblastine, usually bring about only temporary remissions, such drugs can bring about tumour shrinkage, prolongation of life, and significant symptomatic relief.

## MAGNITUDE OF THE CANCER CHALLENGE

Cancer is the second cause of death by a wide margin in the United States. Many more people now survive infectious diseases only to succumb to cancer, and cancer is by no means reserved for the aged. Cancer is either the first or the second cause of death in children between one and fourteen years, with acute leukemia being the most common form of cancer of children. In 1963, 45 per cent of cancer deaths were of persons younger than 65, and 9 per cent were younger than 45 years of age. Cancer of the lung accounts for about one-fourth of all cancer deaths in males.

On the basis of current trends, about one out of every four people alive in the United States today can be expected to develop cancer at some time during his or her lifetime. This means that nearly 50 million people now living in the United States will develop cancer. Also, a little over 30 million Americans now alive will die from cancer unless new natural product chemistry and other endeavours lead to the discovery of new therapeutic procedures for prevention and cure. Economically, all of the costs due to cancer were estimated in 1962 to be 8 billion dollars, or 1·4 per cent of the gross national product.

Louis F. Fieser of Harvard University has made his contributions to the chemistry of natural products. In the April 1966 issue of *Readers Digest*, one may read his own story as to how he narrowly escaped death from lung cancer. Louis states that his case teaches two important lessons: the first one is "the folly of saying it's probably too late to quit now", and the second lesson is that "contrary to my expectations I found it easy to break a long-standing habit of heavy smoking". Of course, he was strongly motivated, but he did stop smoking readily. Suffering also from emphysema, bronchitis, and poor heart function, Louis reported that, after only two weeks, both his heart and lungs showed marked improvement. I commend to you this story of an ex-smoker, and trust that my rather obvious plea will have some benefit.

## AMINO ACIDS OF PRIMARY IMPORTANCE

Nutrition and medicine have benefited enormously from the chemistry of the natural products—amino acids, peptides and proteins. Advances in isolation, automated sequence determination and rapid synthesis of peptide

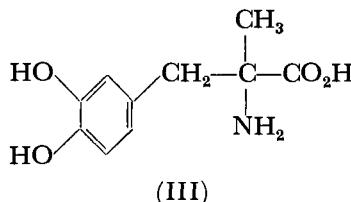
## THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

chains are achievements which might have been viewed as nigh impossible 25 years ago.

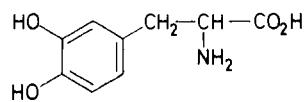
Rather than tell any story about amino acids or peptides, I prefer to cite the story of an amino acid antagonist because of the following three reasons (*i*) this antagonist is extremely useful in current medicine; (*ii*) new antagonists may have future impacts on medicine; and (*iii*) because this story exemplifies the present relationship between a pharmaceutical company and the Food and Drug Administration. About fifteen years ago, Karl Pfister in the Merck Sharp and Dohme Research Laboratories became interested in amino acid antagonists as a possible approach to cancer chemotherapy. Goldenberg of Columbia University, who later died of cardiovascular disease, persuaded Pfister and his group to study the inhibition of the decarboxylation of dopa as a means of possibly controlling essential hypertension.  $\alpha$ -Methyldopa was one of the compounds synthesized and found to be an outstanding inhibitor *in vitro* of pig kidney dopa decarboxylase. However, no drop in the concentration of catecholamine was observed *in vivo*, and, unfortunately, the technique used gave this misleading negative result. The programme on  $\alpha$ -methyldopa all but ended for about five years, and then a paper from Heidelberg reported that  $\alpha$ -methyldopa completely prevented the blood pressure effects produced by dopa in a variety of laboratory animals. Confirmation came from Frankfurt. Udenfriend and Sjoerdsma proposed a reawakened programme on  $\alpha$ -methyldopa between the National Institutes of Health and Merck Sharp and Dohme.

An appreciation of what a pharmaceutical company does these days to gain approval from the Food and Drug Administration of an application for a new synthetic drug may be exemplified by the background on  $\alpha$ -methyldopa for hypertension. In 1959, Udenfriend and Sjoerdsma gave the substance to ten human patients and it showed a reduction in blood pressure confirming the biochemical studies. The rest of 1959 was devoted to chronic toxicity studies and the preclinical phase of Merck and Co., Inc. The clinical studies of 1961 encompassed 200 physicians in the United States and 31 other countries, and nearly 2000 patients. Approval by the Food and Drug Administration was granted in 1962, or about three years from the first successful clinical trial. I have no cost data for this three-year period, but one may surmise that both the costs and the losses would be very impressive. Who else but the pharmaceutical industry would support the cost, the continuity, and provide the extensive skills necessary to carry such a new compound through three trial years so that it could ultimately be marketed to provide hypertensive patients added years of life.

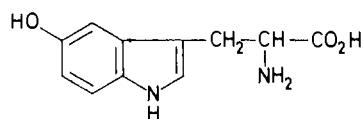
$\alpha$ -Methyldopa [*laevo*-3-(3,4-dihydroxyphenyl)-2-methylalanine] (III) is a decarboxylase inhibitor. The site of inhibition of the biochemical reactions is represented as follows.



KARL FOLKERS

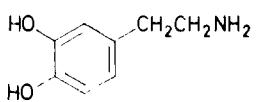


Dihydroxyphenylalanine (dopa)

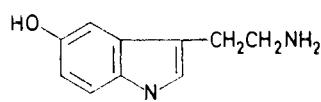


5-Hydroxytryptophan (5-HTP)

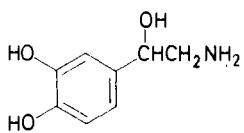
These reactions are inhibited by methyldopa



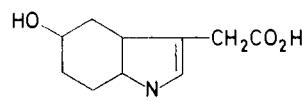
Dihydroxyphenethylamine  
(dopamine)



5-Hydroxytryptamine  
(serotonin)



Norepinephrine



5-Hydroxyindoleacetic acid

Methyldopa finds its greatest usefulness in the field of moderate to severe hypertension and in selected cases of malignant hypertension. One of its chief attributes appears to be the preservation of renal blood flow in patients who have some depression of renal function. It does not generally cause any increase in cardiac rate or output. It produces a smooth and predictable lowering of blood pressure—both the supine and the standing blood pressure.

### **HEART DISEASE AND STROKE—THE GREATEST CHALLENGE**

Surely, new chemistry of natural products will lead to new contributions for the therapy of coronary heart disease, which remains the number one cause of death in the United States. A few decades ago, tuberculosis was the greatest single menace to American health, and pneumonia and influenza took a heavy toll. Parenthetically, may I explain that my use of statistics is for the United States only because of their ready availability to me. Perhaps the corresponding statistics for Europe would not be greatly different.

The term, heart disease, as commonly used, is a complex of diseases and may be better termed as cardiovascular and renal diseases, and these diseases can be divided into three major categories: (i) strokes, or damage to the

## THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

blood vessels affecting the central nervous system; (ii) diseases of the heart and blood vessels, including arteriosclerosis, degenerative heart disease, hypertensive heart disease and other specific disease entities of the heart; and (iii) kidney diseases, including chronic nephritis and renal sclerosis.

These heart diseases accounted for 50 per cent of all deaths in the United States in 1963. Heart diseases are predominantly a cause of death among older people, and men outnumber women as victims by a factor of more than one-third. This complex of heart diseases now claims nearly a million lives each year. The economic impact is also staggering. Had all those individuals who died of heart disease in 1962 lived just one more year, the economy would have gained two billion dollars worth of output. The sum of direct costs, plus losses, amounted to 22·4 billion dollars, or 4 per cent of the gross national product in 1962.

## MALARIA—NOT YET UNDER CONTROL

It is believed that malaria is perhaps the most worldwide of all human diseases, and it was recently estimated that approximately 300 million people are afflicted. Most of the really effective antimalarial drugs presently used in medicine resulted from the intensive investigations before, during, and after World War II in Europe and the U.S.A. There has been a very widespread feeling that malaria is a rapidly disappearing disease. However, the parasitologists tell us that the world control of malaria has not been realized, and that malaria will be a medical problem in the world for a long time. Consequently, those who work together in chemistry and medicine still have research goals in this field. In fact, all the parasitic diseases constitute an enormous challenge for chemistry and medicine. Schistosomiasis, the dysenteries and more than a dozen other parasitic diseases are specific goals.

Drug resistance of malaria became a significant problem about 1948 to 1950 with the discovery in Malaya of the resistance of *Plasmodium falciparum* and *Plasmodium vivax* to Proguanil. Drug resistance to other important antimalarials has been reported from many countries of the world, including Africa and South America.

The war in Vietnam has heightened the interest in drug-resistant strains of malaria and new research programmes have now been launched, including broad basic biological research on experimental malaria in animals, particularly the primate, and finally on intense medical studies of clinical malaria due to resistant strains.

The history of malaria for several thousand years until and including present medical practice demonstrates the longest lasting and most far-reaching contribution of natural products to medicine; because of this extraordinary contribution, and because of the timely focus on resistant malaria in Vietnam, I wish to present a little more of the history and some of the latest findings.

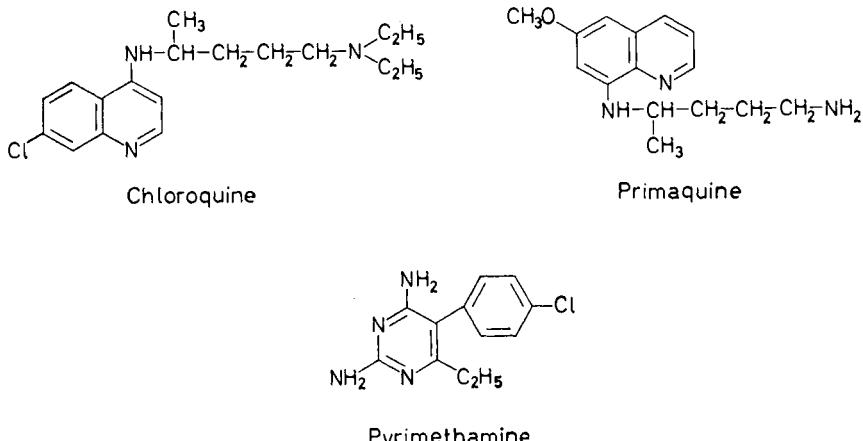
Of all the natural products explored by man to treat malaria, only two are said to have withstood the test of time as well as modern chemistry and medicine. One is the Chinese drug, Ch'ang Shan, which is known to have been used in China for several thousand years. Ch'ang Shan is made from the roots of *Dichroa febrifuga*. Febrifugin has a spectrum of antimalarial

KARL FOLKERS

activity, but its value in human malaria has been limited, because of its side effects.

The other natural product which is really the prime one of the two is, of course, quinine.

Chloroquine, primaquine, amodiaquine, proguanil and pyrimethamine have been regarded as among the best of the available antimalarials. In



mentioning these few synthetic antimalarials, I do so for brevity and to be illustrative, and not to disregard up to twenty other antimalarials which have also had varying degrees of medical success in countries of Africa, Asia and the U.S.S.R., and which stemmed from the world's pharmaceutical industry.

In current research, carefully controlled medical studies, which are quite definitive, have been made on a drug-resistant strain of *Plasmodium falciparum* from both Thailand and Vietnam. The Thailand (JHK) strain<sup>3</sup> and the Vietnam (Sn) strain<sup>4</sup> were both obtained from blood smears of two members of the U.S. Armed Forces in these countries. Infections with both of these strains were established in non-immune volunteers in a non-endemic area under conditions precluding reinfection, that is, a prison in Joliet, Illinois. The volunteers received during acute clinical attacks of malaria chloroquine, hydroxychloroquine, amodiaquine, mepacrine, pyrimethamine, proguanil, or other antimalarials alone and in combination. These drugs failed to effect a radical cure of the infections, both before and after passage of the strain through mosquitos. A radical cure requires complete elimination from the body of erythrocytic stages and persisting tissue stages of the parasite so that relapses cannot occur.

As if in tribute to the activity of quinine, which has served mankind so well for a few hundred years, a radical cure was achieved in these patients by the administration of about 1.5 g of quinine daily for seven days for the resistant *P. falciparum* from Thailand. In the case of the resistant strain from Vietnam, the administration of either 50 mg of pyrimethamine daily for three days or nearly 2 g of quinine daily for ten days did effect radical cures.

The drug resistance of *P. falciparum* to chloroquine is today being

## THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

recognized with increasing frequency, not only in South-East Asia, but in South America. Our medical colleagues, who are very knowledgeable in this field, point out that the present-day global chemotherapy of malaria could be serious.

The prodigious research of chemists on the natural products which have antimalarial activity, and knowledge of the many thousands of organic compounds screened in experimental malaria, lead to the observation that the design of new antimalarials poses formidable new research. Nevertheless, a study of this field, in depth, will reveal new approaches, extensions of old ones, and new discoveries of medical significance could evolve. New fundamental research on parasite metabolism and the biochemistry of active compounds at various stages of the life cycle of the *Plasmodium* may also lead to new and effective drugs. It does seem that there has been little significant progress in the field of malaria chemotherapy in the last ten years as compared with the earlier era.

## 19-NOR-STEROIDS—FERTILITY CONTROL

The ovarian and testicular hormones and those of the corpus luteum were elucidated by chemists working on natural products about 31–37 years ago, and progesterone, which I wish to speak about later, was first isolated in pure form in 1934. It was in 1936 that the alphabet-lettered adrenal steroids reached a peak of chemical elucidation. Concomitantly and independently of this chemistry on the steroid hormones, medical investigators had been extensively studying rheumatoid arthritis. It is difficult to visualize how the chemistry of these steroid hormones and the separate medical investigations would have become bridged to result in the impact of cortisone upon diseases of medicine were it not for the indispensable role of the pharmaceutical industry. In these days of preoccupation with governmental regulations, the pharmaceutical industry of the world must maintain a balance of responsibilities. The history of drugs demonstrates that certain long range investments, which become very large and appear very risky, may turn out to be quite rewarding financially. The industry may find some guidance from the idea that the greater the impact of the chemistry upon medicine the greater will be the impact of the medicine upon economics.

I have particularly mentioned the naturally occurring steroid, progesterone, because I wish to say more about the great current interest in steroid contraceptives. On occasion, these steroids seem to assume the distinction of being known as "the pill". When one considers that by the year 2000 the world's population is projected to become about 7 billion, an increase of 3·5 billion from today, the significance of fertility control is evident. Consequently, there is greatly augmented interest on more basic knowledge of the process of reproduction as well as new progress on all approaches for useful control of conception.

That progesterone inhibited ovulation during pregnancy besides its other biological actions had been known earlier. Ehrenstein synthesized crude "19-nor-progesterone" which was found to exhibit the same biological activity as progesterone. The concept of structural requirements for progesterone activity became more fascinating, and Plattner and Heuser

## KARL FOLKERS

synthesized 14-iso-17-isoprogesterone, but it was found to be inactive. Attention was then focused on synthetic 19-nor-steroids. Although Birch's synthetic 19-nor-testosterone exhibited less androgenic activity than testosterone, a Syntex group synthesized authentic 19-nor-progesterone, because the stereochemistry of Ehrenstein's earlier material was ambiguous. Miramontes, Rosenkranz, and Djerassi reported that authentic 19-nor-progesterone is really 4 to 8 times as active as progesterone in the Clauburg assay.

Before it became established that the potency of progesterone is significantly increased by removal of the methyl group at position 10, it had been observed earlier by Inhoffen, Logemann, Hohlweg, and Serini that 17- $\alpha$ -ethynyltestosterone, by the oral route, was an effective progestational agent. The Syntex group, Djerassi, Miramontes, Rosenkranz, and Sondheimer, then synthesized 19-nor-17- $\alpha$ -ethynyltestosterone. Their anticipation of increased activity by removal of the methyl group at position 10 was realized when the biological data revealed an extremely high order of progestational potency for the compound. Also, it was active by the oral route. 19-Nor-17- $\alpha$ -ethynyltestosterone and its  $\beta,\gamma$ -unsaturated isomer, synthesized by Colten, were the first such steroids studied in human females for inhibition of ovulation, and they were found to be effective by Rock, Pincus, and Garcia. The time between recognition of progesterone as a natural product and 19-nor-17- $\alpha$ -ethynyltestosterone for inhibition of ovulation in humans was 22 years. Again, an indispensable pharmaceutical industry, Syntex, participated in the early stages. Reference to a recent summary<sup>5</sup> by Djerassi on steroid oral contraceptives will provide one with an enlarged account of the chemical developments. On the medical side, the total market in 1965 for the United States for sales of oral contraceptives was estimated at 66.5 million dollars for about 6 million individuals. For 1966, the corresponding sales are estimated to increase to about 85 million dollars for up to 7.5 million individuals.

## VITAMINS AND COENZYMES

The impact of all the known vitamins upon medicine can be summed up in the word *indispensable*. The last vitamin to achieve an established place in medicine is vitamin B<sub>12</sub>, and this was nearly twenty years ago. Before and after B<sub>12</sub>, there has been the question: are there any more vitamins left to discover? Such a question is somewhat imponderable. When Dr. Aulin-Erdtman invited me to make this presentation, she encouraged me to include an account of some of my current research. Consequently, I shall provide some background on coenzyme Q and my progress report from the Stanford Research Institute. Since this research has already been extended to medical data, its inclusion is within the scope of the subject.

Coenzyme Q is a constituent of the mitochondrion and performs its indispensable role of electron transfer in this cytoplasmic organelle. It was as recent as the early fifties that research on the biochemical mechanism of oxidative cycles, electron transport, and oxidative phosphorylation merged with research on cell biology. The mitochondrion of the cell may be visualized as an ellipsoid which is frequently about 3 microns in length and a little less than 1 micron in width. The mitochondrial structure may be represented as

## THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

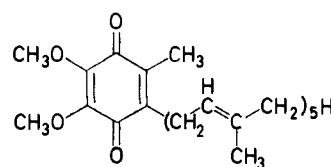
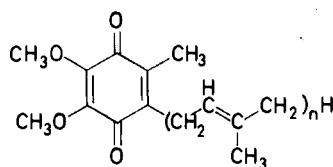
a membrane within a membrane, and the inner membrane has countless infoldings, known as cristae. A schematic representation is given in *Figure 1*.

Mitochondria are relatively free to move in some cells, but those of skeletal muscle appear to remain fixed near contractile macrostructures. The number of mitochondria for a cell varies over a great range. Rat hepatocytes have an average of about 800 mitochondria per cell, and renal tubule cells of mammals contain about 300. Sperm cells can contain as few as 20 mitochondria.

The mitochondrial two-layer membrane system contains two-thirds protein and a noteworthy one-third lipid. About 25 per cent of the total protein of the membrane is composed of the respiratory enzyme assembly. Of great significance, then, is the realization that the mitochondrial membrane bilayer system is not altogether a biochemically inert structure, but includes the recurring multienzyme structure for respiratory electron transfer and the biosynthesis of ATP.

About four years ago, Fernández-Morán made the initial observation that negatively stained images of the inner membrane showed the presence of spheres of about 70–90 Å in diameter which are connected by thin stalks to the surface of the membrane. The existence of these spheres and stalks has been confirmed in other laboratories. An electron micrograph (*Figure 2*), also kindly provided by Dr. Fernández-Morán of the University of Chicago, shows these spheres and stalks. Each mitochondrion appears to contain several thousand of these entities, and they appear to account for about 80 per cent of the dryweight of the mitochondrion. When it is realized that these entities contain complete electron transfer chain(s) and the capacity for coupling, it is apparent that the mitochondrion can be primarily a macrostructure for the electron transfer process of respiration and the biosynthesis of ATP.

The sequence of electron transfer and coupled phosphorylation as depicted by Ernster, Green and others, is shown in *Figure 3*. In this sequence is the expression —CoQ— symbolizing coenzyme Q. In this lecture I am using the coenzyme Q nomenclature in order to emphasize a relationship of this subject to nutrition, although the name ubiquinone is preferred on grounds of nomenclature.

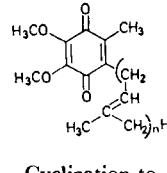


Members of Coenzyme Q group

Coenzyme Q<sub>10</sub> is that member of the CoQ group in the tissue of man, and it has been estimated that the entire human body contains from 1 to 2 grams of it. CoQ<sub>6</sub> through CoQ<sub>10</sub> were isolated in the early days of this research, and very recently coenzyme Q<sub>5</sub> has been found<sup>6</sup> in *E. coli*.

KARL FOLKERS

As shown in *Table 1*, the biological activity of coenzyme Q has certain

<i>Table 1.</i> Organic structural specificity of coenzyme Q for biological activity	
Monoethoxy and diethoxy analogues have low activity	{
Monohydroxy analogue has some inhibitory activity	
Monoamino analogue is naturally occurring rhodoquinone and has low activity	} Isoprenoid chain can be changed in length and somewhat reduced with some retention of activity
	Cyclization to chromenols and chromanols does not inactivate
	Ring methyl group is apparently not too critical for activity

Data given is largely for succinoxidase and partly for DPNH-oxidase system

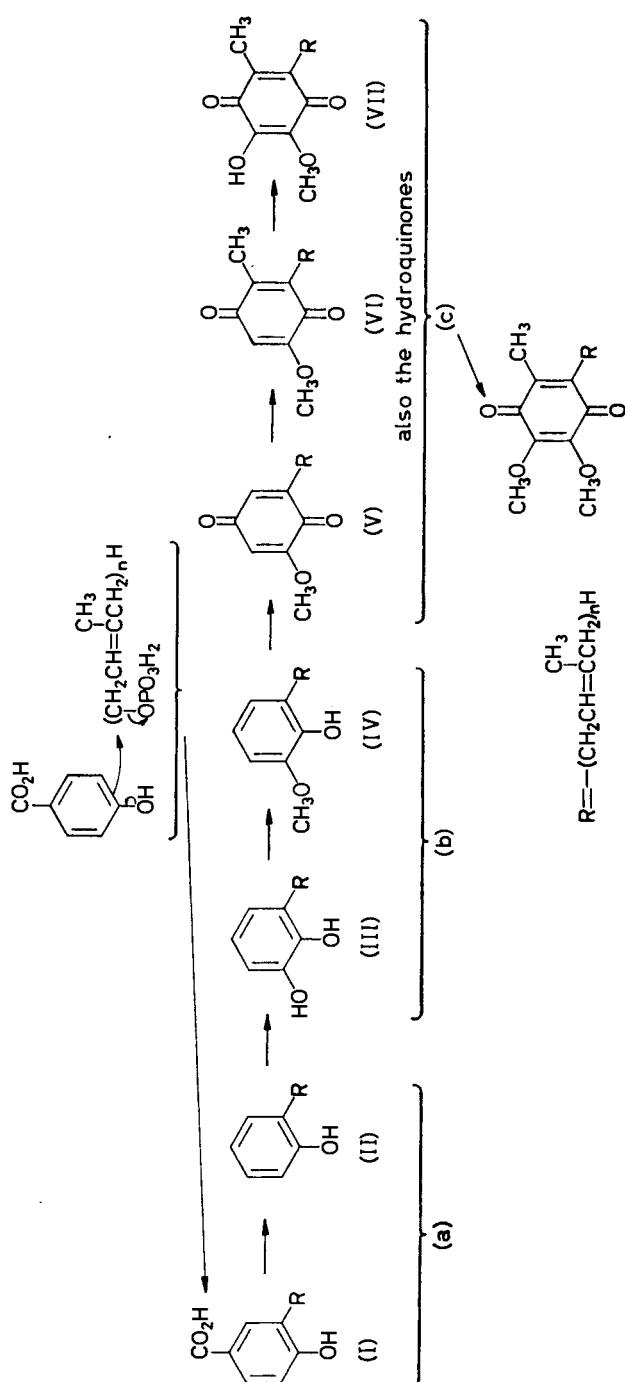
structural specificity in the succinoxidase and NADH-oxidase systems<sup>7</sup>. The variations in structure which permit retention of activity are limited, but are not as rigid as those usually observed for members of the vitamin B complex. The biological activity of coenzyme Q is primarily restricted to compounds having the oxygen and lipidic functionalities of the molecule. Notably, vitamins E and K are inactive in these succinoxidase and NADH-oxidase systems. Consequently, one may consider coenzyme Q in terms of the structural specificity of vitamins for activity.

Coenzyme Q<sub>10</sub> is biosynthesized within the human body, and the sequence of this biosynthesis has challenged a number of investigators for several years. Although it was predictable that the isoprenoid side chain of CoQ would be derived from mevalonic acid and, indeed, this was shown to be true, there was little more than just ideas on how the tetrasubstituted benzoquinone nucleus is biosynthesized. An early understanding that the isoprenoid side chain could be the last step of the biosynthesis now seems to be replaced by evidence that this side chain is actually introduced very early in the sequence. Fortunately, for us who study CoQ in man, the photosynthetic organism, *Rhodospirillum rubrum*, like the human body, utilizes coenzyme Q<sub>10</sub>, and this microorganism has served us well.

*p*-Hydroxybenzoic acid is now established as a precursor to coenzyme Q<sup>8,9</sup>. It has been shown that *p*-hydroxybenzoic acid becomes alkylated and presumably by an isoprenyl pyrophosphate as indicated on the top of *Table 2*. The first two precursors<sup>10</sup> shown within the bracket (a) were elucidated in 1965. The next two precursors<sup>11</sup> within the bracket (b) were elucidated earlier this year. A new report<sup>12</sup> today can now be made on the remaining three steps, all expressed as quinones, prior to the last O-methylation to give coenzyme Q. The quinones (V), (VI), and (VII) are now apparent from the skill, good judgement, and strategic examination of about 5000 fractions from *R. rubrum* by Dr. Palle Friis. Quinone (VI) has been isolated and characterized. Quinone (V) has not yet been isolated in pure form, but spectral data support its presence. The intermediacy of quinone (VII) is obvious. Characterization of quinone (V) is continuing. The intermediacy of the hydroquinones is understood. We believe that the biosynthetic sequence for coenzyme Q in *Table 2* is essentially generic for all members of the coenzyme Q group and all forms of life having CoQ, but

THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

Table 2. Biosynthesis of Coenzyme Q<sub>n</sub>



## KARL FOLKERS

with some variations and alternative pathways, depending upon specific forms of life.

Turning now to animal studies, a pivotal discovery was the finding that hexahydrocoenzyme Q<sub>4</sub> has vitamin-like activity, which is both prophylactic and curative, in the rabbit which becomes dystrophic by nutritional design<sup>13-15</sup> as shown in *Table 3*. Next, this vitamin-like activity was revealed for the anaemic and dystrophic rhesus monkey<sup>16</sup> which is also produced by nutritional design; this activity in the primate has been confirmed and extended<sup>17</sup>, as in *Table 4*. The photograph (*Figure 4*) kindly provided by

*Table 3.* Vitamin-like activity of hexahydrocoenzyme Q<sub>4</sub> in the dystrophic rabbit

Four rabbits are cured of nutritionally induced dystrophy	..	..	..	(a)
Dystrophy is prevented by prophylactic therapy in three rabbits	..	..	..	
Seven rabbits with nutritional dystrophy responded to therapy	..	..	..	(b)
2,3,5-Trimethyl-6-phytylbenzoquinone showed very low activity	..	..	..	
Life-saving activity appraised at dose levels of 2-10 mg/kg	..	..	..	
No evidence found for significant conversion of CoQ <sub>4</sub> -chromanol to CoQ <sub>4</sub>	..	..	..	(c)

*Table 4.* Vitamin-like activity of coenzyme Q in the anaemic and dystrophic rhesus monkey

### Coenzyme Q<sub>10</sub>

Evoked a reticulocytosis in four monkeys  
Slight or no increase in haemoglobin concentration

### Hexahydrocoenzyme Q<sub>4</sub>

Suboptimal dose elicited only a reticulocytosis, but a higher level produced a complete haematological remission in one monkey  
In five monkeys, a haematological response and decrease in creatinurea were clearly produced by therapy

Dr. Coy Fitch is of a dystrophic rhesus monkey. Since the posture is somewhat unusual, it would appear that the monkey is actually dystrophic.

It was next evident that one could forecast a haematological response of children with a unique anaemia to treatment with hexahydrocoenzyme Q<sub>4</sub>. Such children have protein-calorie malnutrition, and a macrocytic anaemia unresponsive to known haematological agents, including folic acid and vitamin B<sub>12</sub>, can develop as first described in 1956-1960 by Amin S. Majaj, Chief of the Paediatrics Department of the Augusta Victoria Hospital, Jerusalem, Jordan<sup>18</sup>.

*Figure 5* shows the picture of such a child (Ibtisam, meaning Smiling) who was recently treated with this CoQ<sub>4</sub> by Dr. Majaj in our continuing co-operation which began in 1963. After four to five days of treatment of this child, the entire haematological status improved very significantly, as evidenced by a complete remission. Also, the creatine excretion of this child decreased after treatment showing improvement in protein metabolism. The data<sup>19</sup> are presented in *Table 5*. After such an encouraging clinical test with CoQ<sub>4</sub>, it can be easily imagined that we were all smiles and are now eagerly planning an extension of the study of refractory anaemia.

During our cooperation in 1963, Dr. Majaj treated two similarly anaemic children with coenzyme Q<sub>10</sub> and observed (unpublished data) a reticulocytosis and improvement in the anaemia; substantiating statements are also

THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

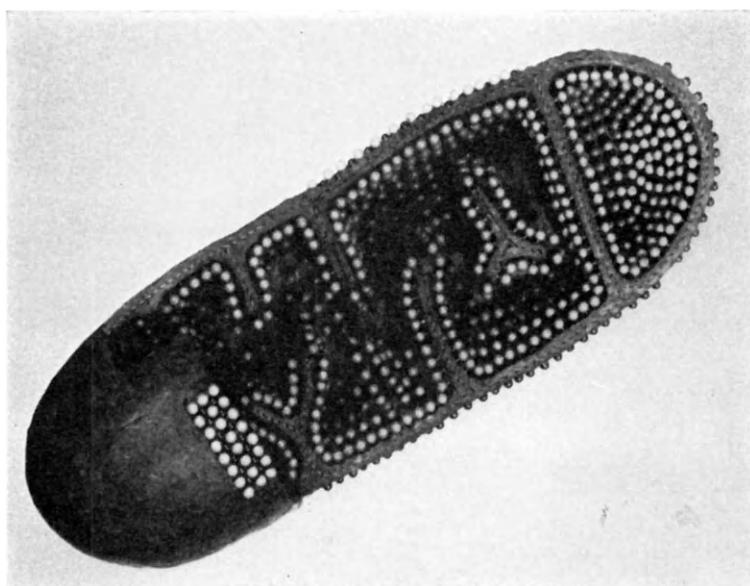


Figure 1. Model of mitochondrion



Figure 2. Electron micrograph of negatively stained mitochondrial membranes

to face p. 14

KARL FOLKERS

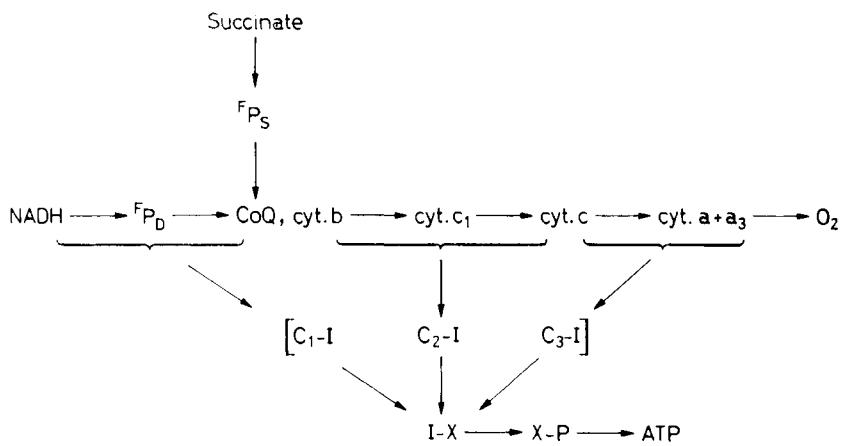
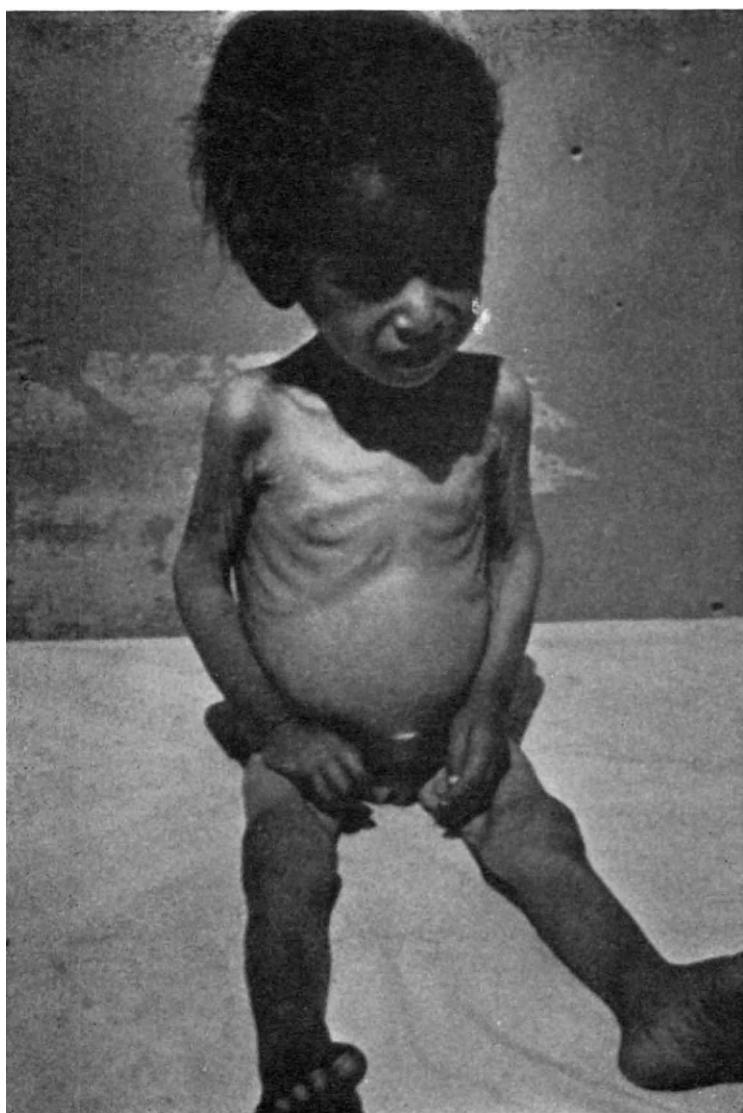


Figure 3. Electron transfer chain



Figure 4. Dystrophic rhesus monkey

THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE



*Figure 5. Child having protein-calorie malnutrition*

KARL FOLKERS



*Figure 6.* Jacobus Berzelius. Portrait by Johan Way 1826. By courtesy of the library of the R. Swedish Academy of Science

## THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE

<i>Table 5.</i> Treatment of a macrocytic anaemia associated with protein-calorie malnutrition	
	<i>With hexahydrocoenzyme Q<sub>4</sub> (first child)</i>
Data on peripheral blood	Reticulocytosis from ca. 0 to 34% Three-fold increase in hematocrit Increase in haemoglobin from 5.5 to 9.5% Increase in leucocytes from 5 100 to 19 350/cm <sup>3</sup>
Data on urine	Decrease towards normalization of creatine excretion
	<i>With coenzyme Q<sub>10</sub> (two children)</i>
Data on blood and marrow	Positive reticulocytosis Change from aplastic to hyperplastic marrow Myeloid: erythroid ratio was 21:1 before treatment and 2:1 after treatment

given in *Table 5*<sup>19</sup>. The fact that these children require more protein and calories does not detract from the significance of their response to therapy with CoQ<sub>4</sub> and CoQ<sub>10</sub>.

Turning again to muscle disease, it has been evident for years that nutritional dystrophy, whatever may be its cause, is clearly metabolically different from genetic dystrophy in mammalian species and particularly in man. Supporting this differentiation is the knowledge that vitamin E is ineffective therapeutically in the treatment of genetically dystrophic mice and muscular dystrophy in man although vitamin E is effective in nutritional dystrophy.

Genetic dystrophy in mice has been known for years, and no significant long-term therapeutic response of intrinsic biochemical significance has yet been reported although, indeed, a variety of treatments have been tried. The statements of basic knowledge which led us to open a study of treatment with CoQ of genetic dystrophy are: (1) CoQ is active in the electron transfer process and coupled biosynthesis of ATP, and vitamin E is inactive; (2) CoQ has exhibited vitamin-like activity in nutritional dystrophy.

Genetically dystrophic mice have now been treated with hexahydrocoenzyme Q<sub>4</sub> and the results are given in *Table 6*<sup>20</sup>. Of eight control animals,

*Table 6.* Response of mice with genetic dystrophy to therapy with hexahydrocoenzyme Q<sub>4</sub>

### *8 Control animals*

Dystrophic status of seven progressively deteriorated  
Dystrophic status of one remained about the same

### *10 Treated animals*

All ten dystrophic mice improved  
Four of the ten were severely dystrophic and in poor health, but responded and were able to walk

the dystrophic status of seven progressively deteriorated. The dystrophic status of one animal remained about the same. Of ten treated animals, all ten dystrophic mice improved. Four of the ten mice were severely dystrophic and in poor health, but even they responded to this therapy and became able to walk using all their legs.

An interpretation of this apparent vitamin-like activity of CoQ in genetic dystrophy may be stated as follows. The biosynthetic sequence for coenzyme Q in mammalian tissue is complex and requires numerous proteins, lipids, known vitamins, inorganic ions, and other cofactors. Such a complex biosynthesis appears to offer many metabolic possibilities for genetic

## KARL FOLKERS

"blocks". From such inborn errors of metabolism, an inadequacy of functional levels of coenzyme Q could result with concomitant impairment of electron transfer and the development of dystrophy. If a deficiency of CoQ actually resulted from such a metabolic "block", then therapeutic administration might well restore CoQ in the tissue to functional levels for electron transfer and diminution of dystrophy. The apparent response of these mice to therapy with this CoQ<sub>4</sub> could signify that this dystrophy of hereditary origin is possibly caused by a genetic "block" in the biosynthesis of CoQ. The long-standing failure of vitamin E to exhibit therapeutic activity in the genetic dystrophy of mice and humans is understandable on the basis that vitamin E should have no effect upon the mechanism of such a genetic "block" itself, and vitamin E cannot substitute for CoQ in electron transfer.

It is now very attractive to extend these favourable data on the therapy of dystrophy of hereditary origin in experimental species and particularly to the muscular dystrophies in humans.

This progress report on CoQ may be concluded with an over-all interpretation as a working basis for new research. It appears that the so-called vitamin E deficiencies can be nonspecific antioxidant deficiencies, primarily because structurally unrelated compounds show the same activity as vitamin E. Such antioxidants could protect, on a body-wide basis, unsaturated lipids, including CoQ and its biosynthetic precursors against peroxidation and loss by a free radical mechanism. In effect, vitamin E could allow functional levels of coenzyme Q to become restored by biosynthesis. There seems to be no proof yet that vitamin E has an intrinsic and specific role in mammalian tissue, like that of a classical vitamin, although such a role may yet be elucidated.

In nutrition, one differentiates between those factors which are just stimulatory from those which are essential to life, and clearly the latter are more significant. Coenzyme Q may have exhibited an essential vitamin-like activity, at least the activity is life-saving, in these studies of the rabbit and the rhesus monkey. The activity of CoQ in these children is probably also life-saving, but one cannot carry the study of these children to such a potentially dangerous end-point. Coenzyme Q may have exhibited a vitamin-like activity in genetic dystrophy by therapeutically restoring functional levels that may be inadequate due to a genetic block of biosynthesis of CoQ.

Functional deficiencies of CoQ could exist in the human body, because of loss through nutritional peroxidation, or through impaired biosynthesis as in malnutrition or through genetic block(s) of biosynthesis.

May I say again that this overall interpretation serves as our present working basis for planning new experiments which I hope will help in the final elucidation of the mechanisms of biological activity.

Hexahydrocoenzyme Q<sub>4</sub> appears to substitute for CoQ<sub>10</sub> in the body.

In conclusion, I wish to quote from Jacobus Berzelius (*Figure 6*), who wrote as follows in his preface addressed to King Gustaf IV Adolf in the first Swedish edition of his *Animal Chemistry* of the year 1806:

"Of all the sciences contributing to medicine, chemistry is the primary one, and, apart from the general light it throws on the entire

THE IMPACT OF NATURAL PRODUCT CHEMISTRY ON MEDICINE  
art of healing, it will soon bestow on some of its branches a perfection  
such as one never could have anticipated."

*Appreciation for support is expressed to Dr. Arnold E. Schaefer, Office of International Research, National Institutes of Health, Bethesda (Md.), Prof. P. A. Plattner, Hoffmann La Roche and Co. Ltd., Basel, Switzerland, Dr. Max Tishler, Merck Sharp and Dohme Research Laboratories, Rahway, N.J.; and the Muscular Dystrophy Associations of America, Inc., New York.*

## References

- <sup>1</sup> Y. Ban and O. Yonemitsu. *Tetrahedron* **20**, 2877 (1964).
- <sup>2</sup> N. Neuss, M. Gorman, W. Hargrove, N. J. Cone, K. Biemann, G. Büchi, and R. E. Manning. *J. Am. Chem. Soc.* **86**, 1440 (1964).
- <sup>3</sup> R. D. Powell, G. J. Brewer, A. S. Alving, and J. W. Millar. *Bull. Wld. Hlth. Org.* **30**, 29 (1964).
- <sup>4</sup> R. D. Powell, G. J. Brewer, R. L. Degowin, and A. S. Alving. *Bull. Wld. Hlth. Org.* **31**, 379 (1964).
- <sup>5</sup> C. Djerassi. *Science* **151**, 1055 (1966).
- <sup>6</sup> P. Friis, G. D. Daves, Jr., and K. Folkers. *Biochem. Biophys. Res. Commun.* **24**, 252 (1966).
- <sup>7</sup> K. Folkers, H. W. Moore, G. Lenaz, and L. Szarkowska. *Biochem. Biophys. Res. Commun.* **23**, 386 (1966).
- <sup>8</sup> W. W. Parson and H. Rudney. *Proc. Natl. Acad. Sci. U.S.* **51**, 444 (1964) and **53**, 599 (1965).
- <sup>9</sup> A. S. Aiyar and R. E. Olson. *Federation Proc.* **23**, 425 (1964).
- <sup>10</sup> R. K. Olsen, J. L. Smith, G. D. Daves, Jr., H. W. Moore, K. Folkers, W. W. Parson, and H. Rudney. *J. Am. Chem. Soc.* **87**, 2298 (1965).
- <sup>11</sup> R. K. Olsen, G. D. Daves, Jr., H. W. Moore, K. Folkers, and H. Rudney. *J. Am. Chem. Soc.* **88**, 2346 (1966).
- <sup>12</sup> P. Friis, G. D. Daves, Jr., and K. Folkers. *J. Am. Chem. Soc.*, **88**, 4754 (1966).
- <sup>13</sup> A. F. Wagner, R. J. Stopkie, and K. Folkers. *Arch. Biochem. Biophys.* **107**, 184 (1964).
- <sup>14</sup> J. L. Smith, H. W. Moore, and K. Folkers. *Proc. Soc. Exptl. Biol. Med.* **118**, 782 (1965).
- <sup>15</sup> J. L. Smith, J. Scholler, H. W. Moore, T. M. Farley, and K. Folkers. *Arch. Biochem. Biophys.* **116**, 129 (1966).
- <sup>16</sup> C. D. Fitch, J. S. Dinning, F. S. Porter, K. Folkers, H. W. Moore, and J. L. Smith. *Arch. Biochem. Biophys.* **112**, 488 (1965).
- <sup>17</sup> T. M. Farley, J. L. Smith, J. Scholler, C. D. Fitch, and K. Folkers. Unpublished data (1966).
- <sup>18</sup> A. Majaj and A. Karakashian. *J. Egypt. M. A.* **39**, 161 (1956); A. Majaj and A. Karakashian. *New England J. Med.* **261**, 359 (1959); A. Majaj. *Gaz. Egypt. Paediat.* **A 8**, 610 (1960).
- <sup>19</sup> A. Majaj, W. J. Rees, and K. Folkers. Manuscript in preparation (1966).
- <sup>20</sup> T. M. Farley, J. Scholler, and K. Folkers. *Biochem. Biophys. Res. Commun.* **24**, 299 (1966).

# **GENERAL METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES**

E. J. COREY

*Harvard University, Cambridge, Massachusetts, U.S.A.*

This paper deals with the subject of synthetic methodology at two very different levels, and accordingly, it is divided into two parts. The first is concerned mainly with the general methods by which the plan of a synthesis is derived, and the second part places emphasis on the operations which constitute the individual steps of a synthesis and on some specific recent developments in this field.

## **PART I**

### **GENERAL PRINCIPLES FOR THE FORMULATION OF A SYNTHETIC PLAN**

The achievement of a synthesis of a complex organic molecule involves a number of distinctly identifiable operations which, however, are not strictly independent of one another. These include the choice of the molecule to be synthesized, the development of a synthetic strategy and plan in general outline, the selection of specific individual steps and their ordering, and the experimental execution of the synthesis. The absence of a clean separation between component elements makes the task of analysing and understanding Synthesis as an intellectual discipline appear quite forbidding, especially in the most crucial aspect, derivation of a general plan. Further, the enormous diversity and number of organic structures now known to exist, the incredibly broad spectrum of reactions available for synthesis, and the uncertain and severe limits on the applicability of any given reaction all combine to create an impression that the design of a synthesis is apt to be tenuously hypothetical and is mainly a function of the unique circumstances in each particular case and, moreover, that considerations of a highly general nature are neither dominant nor very useful. None the less, a sufficiently great number and variety of syntheses have now been completed to encourage attempts at setting down in a generalized form the process by which a synthetic chemist devises an original but valid synthetic route to a complicated structure. Such an effort surely is more than an intriguing theoretical exercise; it is a prerequisite to a deeper comprehension of Synthesis and the methodologies which are fundamental to it, and it is likely to be a keystone in the rational development of Synthesis to still higher forms. For example, any technique for the automatic generation of synthetic schemes by a computer will require a complete and detailed definition of the elements of Synthesis and their mutual interaction, in a most general sense.

### Axioms

There are certain considerations which can be regarded as axiomatic to Synthesis and which serve as pre-conditions for any general analysis of methodology. These include the following:

1. The various elements involved in the solution (and even the selection!) of a synthetic problem are not separable. If a division of these elements in Synthesis is made for purposes of simplifying an analysis, it must be compensated for by allowing their interaction at some stage to produce further modification of the process.

2. A very large number of possible routes to the synthesis of a complex molecule can usually be generated. Each of these involves a sequence of reactions and proceeds via a number of intermediates whose synthesis is more direct than that of the target molecule. Naturally, the starting point for any route should be a readily available synthetic substance.

3. These possible routes are derivable by the recognition of structural units within molecules which can be formed and/or assembled by known or conceivable synthetic operations. (In the discussion which follows, these units are designated as "synthons". In this paper the term "synthetic operations" is used in the molecular sense to denote structural transformations rather than in the laboratory sense, which would imply manipulation.)

4. There are definite but not absolute criteria by which the merits or quality of alternative projected syntheses can be judged. Often there are a sufficiently large number of unknowns to render the selection of a superior synthetic route arbitrary. These criteria, none the less, always serve the very important purpose of dictating the rejection of a very large number of inferior possibilities.

5. The specific criteria by which synthetic alternatives may be judged are on the whole elementary to the organic chemist. The items which follow are important and indicative.

A. The probability of achieving the desired change at each step in the sequence should be high. This implies that at each step the possibility of competing reaction paths should be minimized and that certain controls may have to be introduced in a synthesis to prevent undesired reactions and to position and to orient newly introduced units. The effect of these conditions will be to insure high efficiency in the conversion of input substances to end product.

B. Bypass routes or potential alternatives should exist, particularly where the functioning of one or more of the individual steps is questionable.

C. The solution should be simple. This implies the desirability of a maximum correlation between the individual synthetic operations so that each one permits, assists, or simplifies the others in some way. The absence of correctional steps is also implied.

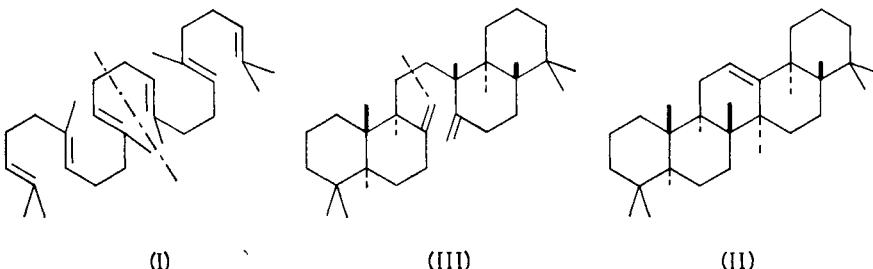
6. If possible, the chemical operations which constitute the individual steps of a synthetic sequence should be chosen from known chemical reactions for which there is a reasonable understanding of mechanism and of scope. However, it may be desirable to attempt a synthesis by the use of new reactions if a great simplification can result or if there is simply no alternative

## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

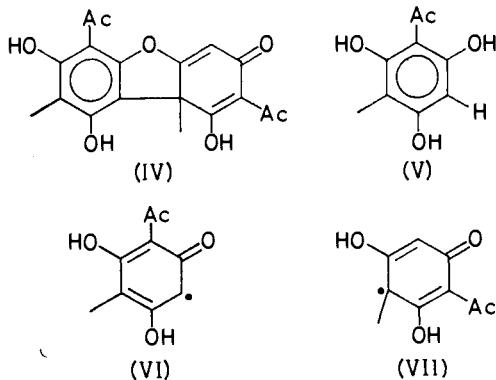
7. Syntheses which can be carried out simply by analogy with closely related known cases are not considered in the category of synthesis under discussion. The synthesis of a new, but simple, indole by a standard method, such as the Fischer synthesis, exemplifies this type of simple synthesis.

### Simplification

Given a complex molecule which is to serve as a target for synthesis, the first step clearly is to attempt a simplification of the problem. Here the analysis of actual or *potential* molecular symmetry can be of importance. The synthesis of squalene (I) is simplified by the fact that the molecule contains two identical  $C_{15}$  units joined symmetrically by a central carbon–carbon bond. Another example of this type of simplification is the synthesis



of the complex alkaloid, C-toxiferine I, from the Wieland–Gumlich aldehyde<sup>1</sup>. In the case of the synthesis  $\gamma$ -onocerane (pentacyclosqualene) (II), the task can be facilitated by the recognition of a synthetic route via the symmetrical intermediate (III); in this case the simplification is less obvious<sup>2</sup>. A slightly different and even more subtle illustration of simplification using symmetry considerations is Barton's synthesis of usnic acid (IV), a molecule which itself appears totally devoid of symmetry but which can be synthesized in one operation by the unsymmetrical coupling of two phenoxy radicals, each derived from the same phenol (V)<sup>3</sup>. These radicals can be formulated as (VI) and (VII). In general, a molecule may be said to possess *potential* symmetry when it can be disconnected to give a symmetrical structure or two or more synthetically equivalent structures.



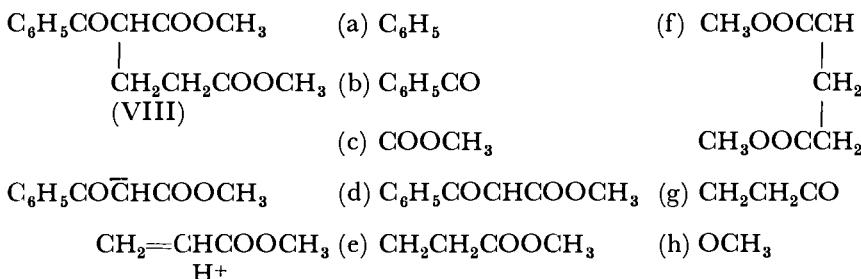
Here a most important point must be made: even in the earliest stages of the process of simplification of a synthetic problem, the chemist must make

### E. J. COREY

use of a particular form of analysis which depends on the interplay between structural features that exist in the target molecule and the types of reactions or synthetic operations available from organic chemistry for the modification or assemblage of structural units. The synthetic chemist has learned by experience to recognize within a target molecule certain units which can be synthesized, modified, or joined by known or conceivable synthetic operations. Thus, with the knowledge of the coupling reactions of phenoxy radicals, enolization reactions, Michael-type addition reactions, and 1,2-elimination reactions, two synthetic units can be recognized within the usnic acid structure which are related to the fragments (VI) and (VII) and, therefore, to the precursor (V).

It is convenient to have a term for such units; the term "synthon" is suggested. These are defined as structural units within a molecule which are related to possible synthetic operations (and, therefore, to the reverse operations of degradation). A synthon may be almost as large as the molecule or as small as a single hydrogen; the same atoms within a molecule may be constituents of several overlapping synthons. Thus, for the molecule (VIII) many synthons can be recognized, including a-h. The units (d) and (e) are valid synthons, since they may be joined by a known synthetic operation, Michael addition (after minor modification to  $C_6H_5CO\bar{C}HCOOCH_3$  and  $CH_2=CH-COOCH_3 + H^+$ )

#### SYNTHONS



In general, the greater the number and variety of synthons which can be defined, the greater will be the complexity of the synthetic problem. Further, recognition of some synthons (usually the larger or major synthons) normally is more useful in analysis than that of others. The greater the degree of connectivity within a molecule (*e.g.*, the larger the number of rings), the greater the number of possible synthons. In principle there is no reason why the number of derivable synthons cannot exceed the number of atoms in a molecule rich in internal connection.

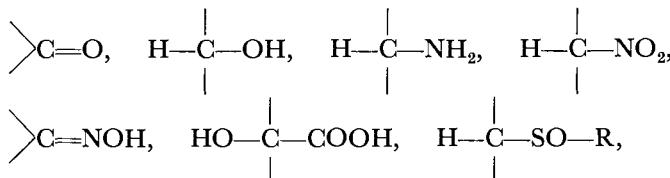
The recognition of synthons within molecules is purely a utilitarian device; the derivation of all the possible synthons in a molecule may never be required. On the other hand, the consideration of a molecule as a collection of the constituent atoms is perfectly definite, but hardly useful in the design of a synthesis.

The beginning student of organic chemistry, when asked to devise a synthesis, such as the formation of butane-1,3-diol from a 2-carbon precursor, often does not appreciate the beauty of the problem, at least in part because

## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

he has generally not been instructed in the simple but fundamental recognition process on which is based all synthetic design, from the simplest to the most complex.

A functional group may be regarded as a synthon, and it is helpful to consider sets of equivalent synthons, *e.g.*,

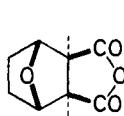


which may be interconvertible by synthetic operations. In some cases a synthetic problem is simplified by replacing one or more synthons by equivalent ones. Clearly, this procedure can also lead to a reduction in the number of functional groups: for example, one C=C might replace two carbonyl groups.

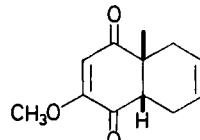
The *next* stage of the simplification process often derives from a consideration of the three-dimensional character (*i.e.*, stereostructure) of a molecule, more particularly, with reference to the following features: centres of asymmetry, orientational and conformational stability, and proximity of pairs or sets of groups.

With regard to asymmetric centres it is apparent that the configuration of each individual centre must be taken into account, as must the influence of these centres on one another. In the extreme case in which the asymmetric centres are distant and effectively isolated from one another, the problem of stereospecific synthesis can be difficult. Possible measures for the simplification of such problems include the use of control units which allow a pre-existing centre to influence the formation of another, the use of dissymmetric reagents or catalysts, and the joining of structural units in which *all* the asymmetric centres are *preformed* with the proper absolute configuration. The last technique is illustrated by the standard synthesis of poly-peptides from amino acids of a specified configuration.

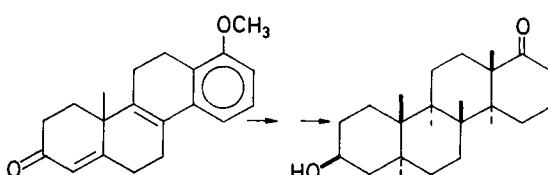
Molecules in which centres of asymmetry effectively interact present different challenges and new opportunities for simplification. Often in these cases the synthetic scheme may depend greatly on the stereoisomer to be



(IX)



(X)



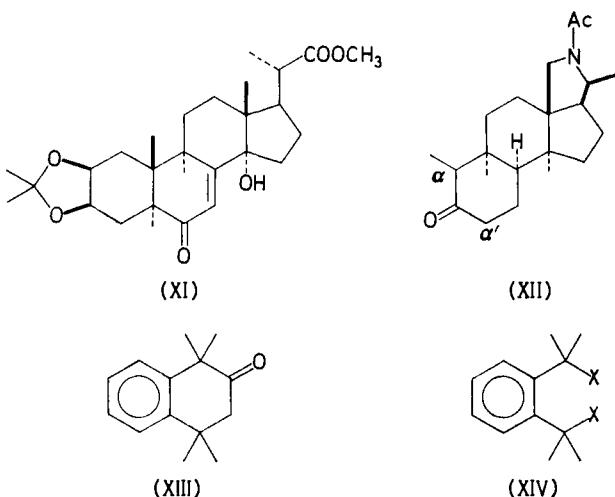
E. J. COREY

constructed. The Stork synthesis of cantharidin (**IX**) is a good illustration<sup>4</sup>.

When asymmetric centres interact, the various possible stereoisomers and the transition states leading to them are of different energy. Relative stabilities then depend heavily on spatial relationships. The application of conformational analysis, the distinction between stable and unstable orientations, and the possible use of processes which equilibrate isomers all become of great importance. A stable arrangement can often be produced by processes of thermodynamic or equilibrium control as well as kinetic control. A less stable arrangement must usually (but not always) be constructed by a stereoselective, kinetically controlled process.

There are widely applicable techniques for the establishment of asymmetric centres in the less stable arrangement which depend on proximity or steric effects. Such processes, *e.g.*, the borohydride reduction of an 11-keto steroid to give the 11 $\beta$ -alcohol, are often described as involving attack by a reagent from the less hindered approach. By means of such reactions proximity effects can be self-amplifying and can influence more strongly the creation of further centres of asymmetry.

Several recent syntheses of decalin systems provide clear illustrations of these general statements. Many of the existing syntheses of steroids have taken advantage of the fact that the ring fusions are in the thermodynamically stable *trans* arrangement. In the Woodward total synthesis<sup>5</sup> the *cis* fusion, established initially in the first precursor (**X**) of the C/D ring system by a Diels-Alder reaction, was later modified to the desired *trans*-locked ring system by equilibration via the keto $\rightleftharpoons$ enol change. The synthesis of *epi*-androsterone by W. S. Johnson<sup>6</sup> was greatly simplified by the *anti-trans-anti-trans* geometry, which allowed the efficient introduction of six centres of asymmetry in two steps. On the other hand, synthesis of the *cis* A/B fusion of the insect hormone ecdysone was accomplished via the intermediate (**XI**), which has *trans*-locked A/B rings, by an enol $\rightleftharpoons$ keto equilibration, with the 2 $\beta$ ,3 $\beta$ -acetonide synthon serving as controller to reverse the normal *cis-trans* energy relationship<sup>7</sup>.

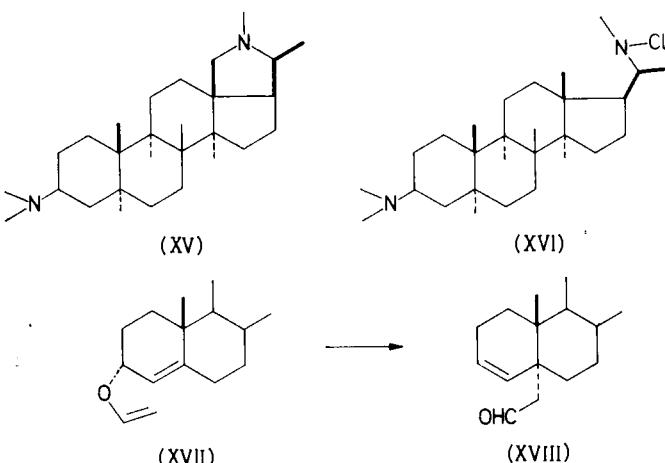


## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

It may at times be desirable to design the synthesis of a thermodynamically stable structure via an intermediate with a different and relatively unstable geometry. Commonly, this indirectness is a price which is paid when the Diels-Alder reaction is applied to fused ring formation. Another interesting example occurs in Stork's synthesis of conessine<sup>8</sup>, in which the intermediate (XII) was deliberately created (by *cis* catalytic hydrogenation) in the unstable, B/C-*cis* fusion, even though the B/C fusion of conessine is the stable *trans* form, in order to direct enolate formation in (XII) to C<sub>α</sub> rather than to C<sub>α'</sub>. Of course, the correct B/C fusion was subsequently obtained by equilibration.

The spatial proximity of groups within a molecule can be of enormous consequence to synthesis. In 1,2-di-*t*-butylbenzene, for example, the proximity of the two *t*-butyl groups leads to severe destabilization and renders the classical approach for the direct introduction of a *t*-butyl group on a benzene ring unworkable. The same proximity has allowed the uncomplicated synthesis from the bicyclic structure (XIII) via (XIV; X = COOH), and (XIV; X = CH<sub>2</sub>OTs)<sup>9</sup>.

When an analysis of a stereostructure leads to the conclusion that strong repulsive, non-bonded interactions are present, it can be expected that any synthetic scheme which does not take full cognizance of stereochemical possibilities or problems will have a correspondingly low expectation of success. At the opposite end of the scale are cases such as adamantane, a structure of such stability that it can be reached from innumerable saturated hydrocarbons by equilibrating isomerization<sup>10</sup>.



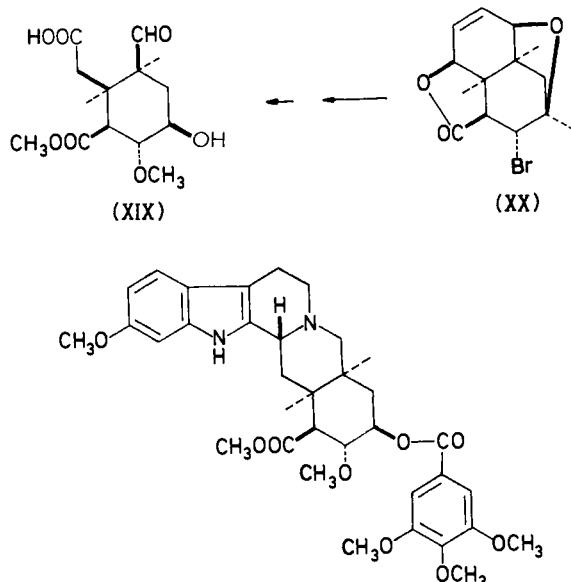
A somewhat more subtle use of proximity was involved in the synthesis of dihydroconessine (XV) from a pregnane derivative without a functional group at C<sub>18</sub>. The proximity of the nitrogen substituent at C<sub>21</sub> and the angular methyl C<sub>18</sub> in the synthetic intermediate (XVI) directs selective free-radical attack on hydrogen at C<sub>18</sub> to produce first an 18-chloro-20-amino compound and then dihydroconessine<sup>11</sup>. This functionalization process has subsequently been extended greatly. The reaction of alcohols with lead tetraacetate<sup>12</sup>, and the photolysis of hypohalites<sup>13</sup> and nitrates<sup>14</sup> are now

E. J. COREY

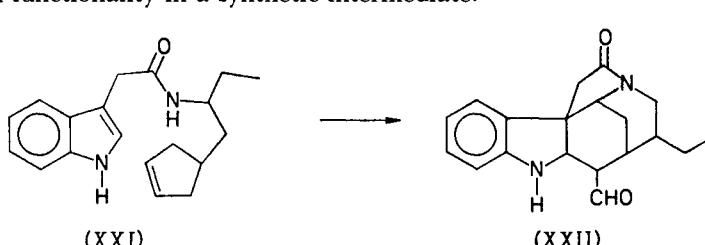
extremely important techniques of synthesis which depend on proximity effects.

A variety of examples could be given to illustrate the use of positioned functionality for the delivery of groups to specific *reactive* centres; the charming application of the Claisen rearrangement by Burgstahler (XVII $\rightarrow$ XVIII) belongs in this category<sup>15</sup>.

Proximity of *reactive* functional groups is of major importance in synthesis. The interactions which result from such proximity give rise to the phenomenon of "neighbouring groups", in which a reaction is channelled along a course involving temporary rings in cyclic intermediates with a major effect on rate, stereochemistry, and product distribution. Spatial proximity of functional groups obviously influences the possibilities for the use of control and equivalent synthons. Indeed, this is now almost a synthetic cliché. The Woodward synthesis of the monocyclic reserpine intermediate (XIX) via the tetracyclic structure (XX) is an instructive example<sup>16</sup>.

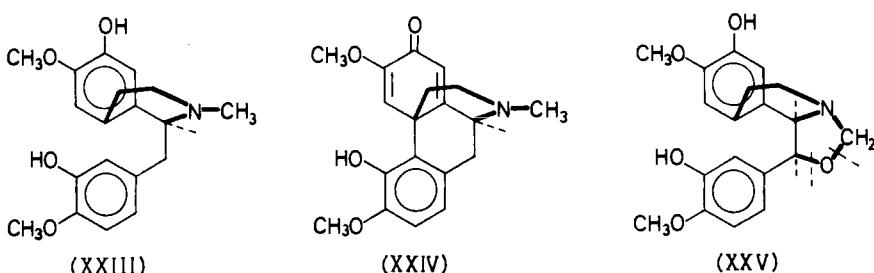


The two-step synthesis (XXI  $\rightarrow$  XXII) of a strychnine-type structure by van Tamelen<sup>17</sup> represents an elegant application of rings for the minimization of functionality in a synthetic intermediate.



## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

At a slightly different level these arguments lead to the proposition that in certain synthetic problems it may be extremely helpful to induce proximity by the use of temporary bridging elements as control synthons. Such considerations suggest, for example, that the phenol coupling process (**XXIII** → **XXIV**) which provides a synthetic route to morphine and which has been found by Barton<sup>18</sup> to proceed in only 0·012 per cent yield, could better be carried out with an intermediate such as (**XXV**) in which the molecule is held in a geometry suitable for directed cyclization. It is obvious that such a synthetic technique is an intramolecular version of the intermolecular control situation which is central to the functioning of enzymatic catalysts in the synthesis of complex molecules.



The use of metal ions or atoms as directing elements in synthesis is particularly intriguing. The original synthesis of phthalocyanines<sup>19</sup>, the syntheses of macrocycles by Busch<sup>20</sup>, the synthesis of cyclododecatriene by Wilke<sup>21</sup>, and the brilliant work of Eschenmoser<sup>22</sup> on the synthesis of corrins are illustrative of the possibilities in this area.

Knowledge or enlightened speculation as to the way in which a molecule to be synthesized has previously been brought into existence can be extremely helpful in simplifying a synthetic problem. In the case of natural products, the mode of biosynthesis is relevant; in the synthesis of usnic acid (IV), referred to above, a knowledge of biosynthesis played a role in suggesting the simplified solution. In the field of alkaloid synthesis the use of biogenetic considerations has been very fruitful from the time of Robinson's pioneering synthesis of tropinone<sup>23</sup> to the recent scheme of Battersby for colchicine<sup>24</sup>. A simplification in the synthesis of  $\alpha$ -caryophyllene alcohol came from a speculation as to the mechanism of its origin from humulene<sup>25</sup>.

The stage of molecular simplification is not really separable from those which are described below. *Some of the most important simplifications in a synthetic problem can be uncovered only during the process of deriving a sequence of synthetic intermediates.* An operational loop must be established which allows interaction of the initial ideas for simplification with those which develop at the later stages of analysis of the problem.

When simplification has been carried to the limit, the next phase of analysis is reached—generation of further sets of intermediates. Simplification may in some instances have proceeded far enough to establish the form of a simple overall plan. Clearly, additional simplification must be sought for each successive intermediate in progressing toward the starting point for synthesis.

## **Generation of sequences and specific intermediates**

The task of devising a definite set of synthetic intermediates connected by specific synthetic operations is usually simplest for acyclic structures. Within this category, as is generally the case, a number of factors can complicate the problem, including substitution and functionality, elements of instability, and the presence of asymmetric centres. The joining of cyclic units to such an acyclic molecule usually does not provide a major complication in cases where the construction of the cyclic units themselves is trivial, *e.g.*, benzene or other standard ring systems. However, structures containing ring(s) with features not readily introduced by standard synthetic operations (the result, *e.g.*, of ring size, strain, or functionality) often present a great challenge. Furthermore, polycyclic structures which have a high degree of internal connectivity (many common atoms) present still more formidable obstacles, usually (but not always) in proportion to internal connectivity as well as substitution, functionality and sources of asymmetry. It is advantageous to consider first the general principles behind the generation of a sequence of synthetic intermediates for acyclic cases and then to examine the modifications or elaborations which may be appropriate to cyclic structures.

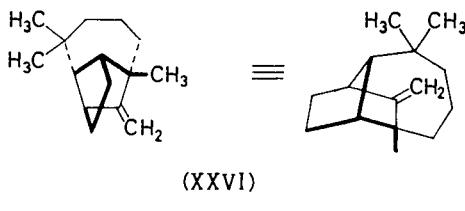
For acyclic molecules the recognition of important synthons is a prerequisite to the generation of possible synthetic intermediates. In the case of highly repetitive structures, *e.g.*, peptides, this exercise is so simple as to be obvious. In more highly variegated structures the recognition of synthons is both difficult and more important. Here smaller molecular fragments are generated by disconnection of synthons, either directly, after the introduction of equivalent synthons, the introduction of control synthons, or the introduction of rearranged synthons. The synthesis of methyl *t*-butyl ketone (pinacolone) provides a simple example of the value of considering rearrangement; the synthesis of 1,3-butanediol shows the value of considering equivalent synthons, since this leads to the intermediates  $\beta$ -hydroxy *n*-butyraldehyde and acetaldehyde. The use of cyclic intermediates to produce difunctionality is another form of the generation of equivalent synthons, *e.g.*, the change of a cyclohexene to a  $\Delta^1$ -cyclopentenecarboxaldehyde via the 1,6-hexanodial system. Protecting groups and temporary rings are useful control synthons. The latter often serve to facilitate the introduction of functional groups, molecular fragments, or centres of asymmetry. From these considerations, and others which need no elaboration, it can be seen that the generation of a set of synthetic intermediates for an acyclic target molecule and the design of a synthesis include the following:

1. Simplification of problem.
2. Systematic recognition of synthons.
3. Generation of equivalent and modified synthons.
4. Addition of control synthons.
5. Systematic disconnection of synthons.
6. Formulation of the possible synthetic transformations which reform the starting structure from the derived intermediate(s).
7. Repetition of items 1–6 for each intermediate and each sequence (parallel sequences may be generated), including previously generated intermediates.

## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

8. Generation of intermediates until the required starting point is reached.
9. Removal of inconsistencies.
10. Identification of unresolved problems.
11. Repetition of items 1–10 to generate alternative schemes.
12. Assignment of merit.

The same steps are involved in the formulation of a synthetic scheme for cyclic molecules, but some important considerations must be added. The disconnection of synthons in cyclic structures may involve the breaking of one, two or perhaps three bonds; the effect may be to break rings without the formation of fragments, to break rings with the formation of fragments or to form fragments without breaking rings. In general, more steps of disconnection will be required to arrive at simpler structures from polycyclic molecules of high internal connectivity. Further, the number of possible synthons is large in relation to molecular size, and the number of possible synthetic designs which can be devised is correspondingly great. The difficulties can obviously be compounded by the presence of exotic rings, substitution patterns or functionality. The number of synthetic schemes which can be derived for the steroid system or even for cyclic C<sub>15</sub> terpenoids is impressive. Formal procedures for the systematic disconnection of the molecular network and the ordering of a synthetic analysis become a requirement. An example of such a procedure is to be found in the published work on the synthesis of longifolene (XXVI)<sup>26</sup>.



The list of items 1–12 which comprise the abbreviated directions for devising a synthesis really describes a cyclic iterative process. It should be emphasized that these directions imply that whenever a new intermediate is generated, the previously derived intermediates must be subjected to possible modifications in response to requirements which develop at a later stage in the analysis. Gradually a more self-consistent and more valid sequence should result. It also should be noted that reasonable alternative syntheses often can be formulated simply by rearranging the sequence of individual steps.

This list of items 1–12 in the set of general instructions given above bears a vague resemblance to a computer programme; its resemblance to processes by which many of the well-known syntheses of today have been conceived is presumably much better, although many more steps of analysis are implied by 1–12 than have customarily been used. Furthermore, in many successful syntheses, one can be assured, not all of the intermediates were specified in detail beforehand, if only because the chemical situations involved in the syntheses were at some stage too complex to allow clear

### E. J. COREY

predictions to be made. In such cases the general outline of a synthesis serves to lend a sense of direction, and the assumption is made that the experimental results will illuminate the fine detail sufficiently to guide the synthesis through the region of uncertainty, a situation not unlike the process of climbing a mountain or traversing a wilderness without benefit of map or trail. Some of the greatest syntheses certainly were accomplished by skilled practitioners working in this venturesome vein. It is probably safe to assume that in all syntheses there is some interaction of this character between the elements of planning and of experimental execution.

The synthetic chemist is more than a logician and strategist; he is an explorer strongly influenced to speculate, to imagine, and even to create. These added elements provide the touch of artistry which can hardly be included in a cataloguing of the basic principles of Synthesis, but they are very real and extremely important. Further, it must be emphasized that intellectual processes such as the recognition and use of synthons require considerable ability and knowledge; here, too, genius and originality find ample opportunity for expression.

The proposition can be advanced that many of the most distinguished synthetic studies have entailed a balance between two different research philosophies, one embodying the ideal of a deductive analysis based on known methodology and current theory, and the other emphasizing innovation and even speculation. The appeal of a problem in synthesis and its attractiveness can be expected to reach a level out of all proportion to practical considerations whenever it presents a clear challenge to the creativity, originality and imagination of the expert in synthesis.

## PART II

### METHODOLOGY OF INDIVIDUAL STEPS

The topic which is bounded by the title of this section is essentially Organic Chemistry in *all its forms*, including known reactions, *reaction theory*, *stereochemistry* and *experimental practice* as well as the physical and analytical aspects of organic chemistry. It is most important for purposes of this essay to focus on synthetic operations and reactions. In general, the following basic types of synthetic transformations can be recognized:

1. Modification of chains: lengthening, shortening, branching, rearranging.
2. Formation of rings.
3. Modification of rings: cleavage contraction, expansion, rearrangement.
4. Introduction of functional groups.
5. Modification of functional groups: removal, interconversion.
6. Control operations: introduction, utilization, removal of protecting and directing groups.
7. Modification of stereochemistry.

Organic reactions and processes can be classified according to their place within the above list of operations just as validly as they can be catalogued according to mechanism. In fact, *both* classifications are necessary to

## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

Synthesis. It is clear, in addition, that a knowledge of the selectivity, scope and stereochemistry of reactions also plays a vital part in Synthesis.

Both the analysis of a synthetic problem to produce a plan of Synthesis and the actual execution of the Synthesis depend critically on the methods available for transforming one organic substance into another. As new methods are developed, both tasks are simplified. Furthermore, the addition of new methods of Synthesis which are general and which, therefore, can be used broadly effectively multiplies the utility of previously existing methods. The importance of developing new general methods which are applicable to the synthesis of complex molecules is made even more apparent by examination of the syntheses which have been completed during the span 1945–1966, the period which encompasses most of the major achievements of Synthesis.

Almost all of the salient syntheses of this time have depended heavily on one or more synthetic methods developed subsequent to 1940. The majority of these new synthetic methods were uncovered by the systematic study of the fundamental chemistry of certain classes of organic compounds. However, new methods are now being developed at an ever increasing rate by a different approach; one in which certain synthetic operations, not allowed by pre-existing reactions, become the primary goal of research. One specific example is the search for methods for the introduction of functional groups at unactivated sites in molecules. The discovery of new methods for performing specific structural operations requires first a *realization* of certain unfilled needs in the field of Synthesis. The combination of reaction theory with a broad appreciation of classical and contemporary chemistry and a willingness to speculate and experiment is also vital.

### SOME NEW SYNTHETIC METHODS

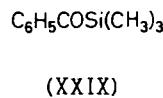
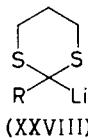
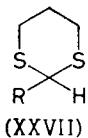
In the final section of this paper, a few of the new synthetic methods which we have recently developed will be described briefly.

#### **1. Symmetrization of reactivity; nucleophilic carbon from carbonyl groups**

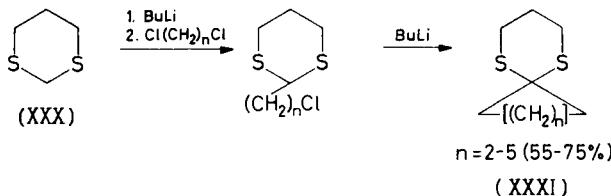
A general class of synthetic operations whose development holds great future promise is composed of processes which in some way *reverse temporarily* the characteristic type of reactivity, nucleophilic or electrophilic, of an atom in a functional group. We have described one such process which allows transformation of the normally electrophilic carbonyl carbon of aldehydes into a nucleophilic centre<sup>27</sup>. The carbonyl group is converted into a 1,3-dithiane system (XXVII), which can be metalated easily by *n*-butyllithium at C<sub>2</sub>. The resulting lithio derivatives (XXVIII) undergo the whole gamut of reactions characteristic of organolithium compounds; they react with carbonyl groups and other electrophilic double bonds, with primary and secondary halides, epoxides, etc. Conversion of the 1,3-dithianes after such transformations to the corresponding carbonyl compound can be carried out under catalysis by acids or mercuric ion, or oxidatively with N-bromo-succinimide under mild conditions<sup>28</sup>. The use of these reagents has been applied to the synthesis of aldehydes, ketones, carboxylic acids,  $\alpha$ -diketones,

E. J. COREY

$\alpha$ -hydroxyketones,  $\alpha$ -aminoketones, 1,*n*-diketones,  $\alpha$ -keto acids,  $\beta$ -hydroxyketones,  $\alpha,\beta$ -unsaturated ketones, and even silyl ketones such as (XXIX).



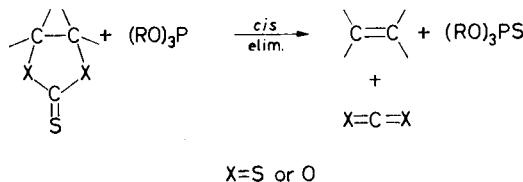
New methods for the efficient formation of rings from 1,3-dithianes have also been developed, as illustrated, for example, by the synthesis of a series of thioketals of cyclic ketones (XXXI) from 1,3-dithiane (XXX) and 1,*n*-dihalides.



The anions derived from 1,3-dithianes represent powerful synthetic tools which in the future are likely to find frequent use in the synthesis of complex molecules. These reagents have already been applied successfully in a number of syntheses of natural products now being carried out in our laboratories, and our experiences thus far lead us to urge that these new synthetic reagents not be overlooked by those engaged in such work. The utility of the 1,3-dithiane reagents provides additional incentive for further investigations towards the discovery of new methods for symmetrizing reactivity relationships in synthesis.

## 2. Synthesis of olefins by stereospecific elimination

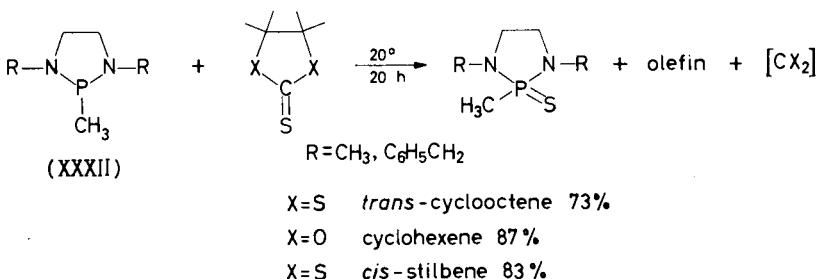
A new stereospecific and position-specific synthesis of olefins has recently been described<sup>29</sup> which follows the general scheme:



This process proceeds with complete stereospecificity by a *cis*-elimination pathway. It allows the stereospecific synthesis of strained cycloolefins, e.g., *cis* or *trans* cyclooctene, the interconversion of *cis* and *trans* olefins, and even the generation of extremely unstable structures, e.g., *trans*-cycloheptene<sup>30</sup>.

Until recently trialkyl phosphites were the most effective reagents known for this synthesis, and they were used at 100–130°. However, the diaza-phospholidine (XXXII) has now been found to be more reactive by several

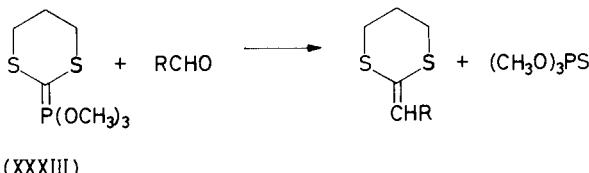
METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES



orders of magnitude and to permit the generation of olefins at room temperature<sup>31</sup>.

The mildness of these conditions should allow the extension of the new olefin synthesis to a wide range of complex and/or unstable structures.

The new elimination method cannot be extended to 1,3- or  $\gamma$ -elimination reactions to form cyclopropanes. In the case of trimethylene-1,3-trithiocarbonate, reaction with trimethyl phosphite produces the phosphite ylide (XXXIII), which shows no tendency to decompose to cyclopropane. This ylide reacts smoothly, however, with aldehydes (but not with ketones) by a Wittig pathway<sup>32</sup>:

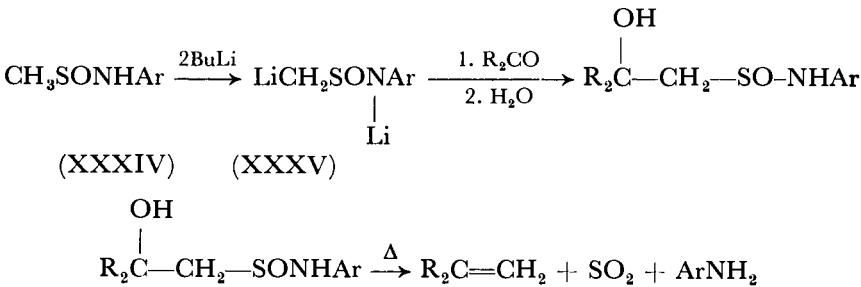


In effect this reaction provides a method for the conversion of RCHO to RCH<sub>2</sub>COOH which is extremely selective.

### 3. Synthesis of olefins by coupling and stereospecific elimination

New olefin syntheses have also been developed from  $\alpha$ -lithio sulfinamides and  $\alpha$ -lithio phosphonamides. These involve the same overall conversion of a carbonyl to an olefinic function which is effected in the Wittig reaction, but there are important and useful differences.

Reaction of methanesulfinic acid *p*-toluidide (XXXIV) with two equivalents of *n*-butyllithium ( $-78^\circ$ , THF) produces the dilithio derivative (XXXV)<sup>33</sup>. This reagent combines with ketones to form easily purified crystalline  $\beta$ -hydroxysulfinamides which undergo decomposition upon

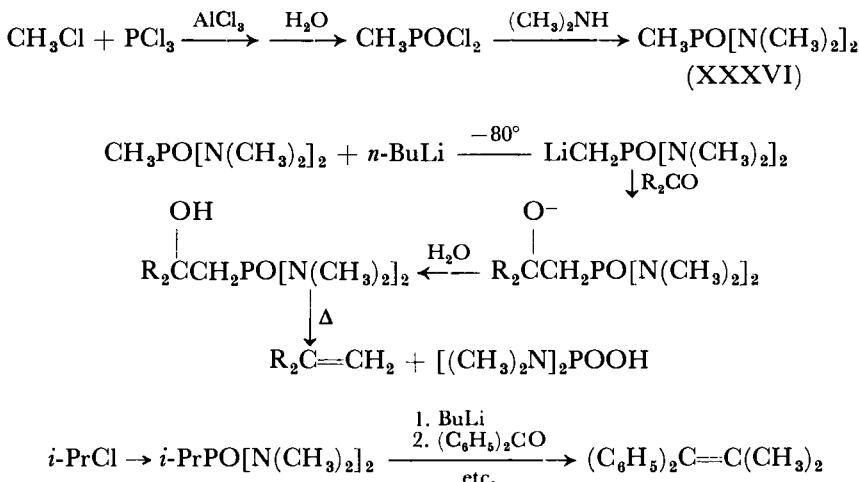


E. J. COREY

heating at reflux in benzene to form olefins, sulfur dioxide, and *p*-toluidine. It is noteworthy that  $\beta$ -alkoxysulfonamide anions do not tend to decompose to olefins. Carbonyl adducts can also be obtained from aldehydes, but their conversion to olefins is rather less efficient than that observed with adducts from ketones; in general, the ease and efficiency of olefin formation increases with the degree of substitution at the olefinic bond.

Reaction of alkylphosphonic acid bis-dimethylamides with *n*-butyllithium ( $-78^\circ$ , THF) leads generally to the  $\alpha$ -lithio derivatives which react with aldehydes and ketones to form easily purified, crystalline  $\beta$ -hydroxyphosphonic acid amides in high yields.  $\beta$ -Hydroxyphosphonic acid amides undergo elimination smoothly upon heating at reflux in benzene to form olefins. These general reactions are illustrated for the specific case of methylphosphonic acid bis-dimethylamide (XXXVI).

Phosphonamide Route to Olefins



Certain potential advantages of the phosphonamide route to olefins as compared with the Wittig reaction can be cited. (1) The reagents are readily available and potentially much cheaper, especially on a molar basis. (2) The elimination of  $\beta$ -hydroxyphosphonic acid amides to form *cis* or *trans* olefins is stereospecific. Therefore, by controlling or modifying the configuration of the intermediate, either a *cis* or *trans* olefin can be generated. Thus, this olefin synthesis also allows control over both geometry and position of the ethylenic unit. The synthesis of *cis*- and *trans*-1-phenylpropene can be carried out stereospecifically by elimination from the diastereomeric adducts (XXXVII). (3) The  $\alpha$ -lithio phosphonic acid amides can be alkylated efficiently, in contrast to phosphonium ylids as is shown by the formation of (XXXVIII).

Experimental data on the advantages of phosphonamide route to olefins are provided in *Tables 1* and *2*.

## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

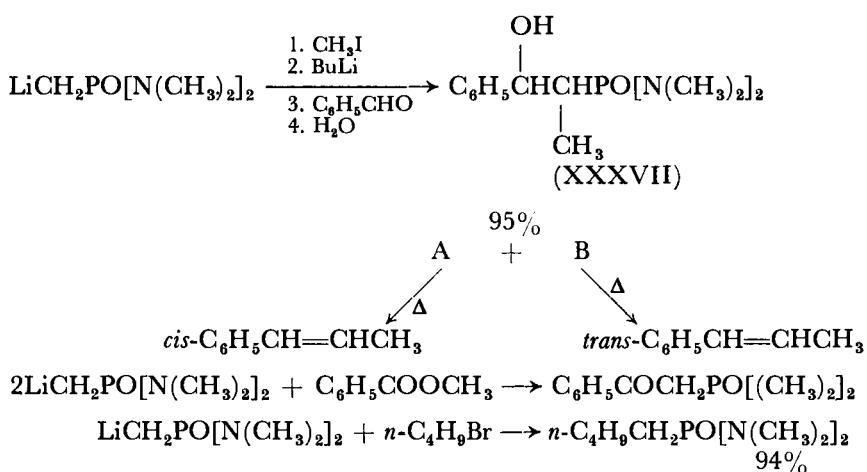


Table 1. Conversion  $\text{R}_1\text{R}_2\text{C=O} \rightarrow \text{R}_1\text{R}_2\text{C=CH}_2$  via methylphosphonic acid (Bis-dimethylamide adducts)

Carbonyl Compound	Yield of adduct (%)	Yield of olefin (%)
Benzophenone	95	93
4-t-Butylcyclohexanone	98	65
2-Cyclohexenone	96	78
Benzaldehyde	95	53
$\Delta^3$ -Cyclohexenecarboxaldehyde	95	67
Dodecanal	89	70

Table 2. Conversion  $\text{R}_1\text{R}_2\text{CO} \rightarrow \text{R}_1\text{R}_2\text{C=CHCH}_3$  via ethylphosphonic acid (Bis-dimethylamide adducts)

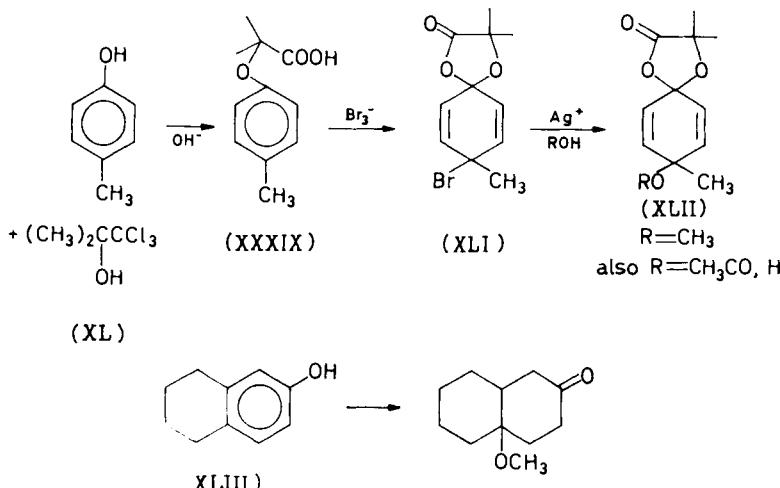
Carbonyl Compound	Yield of adduct (%)	Yield of olefin (%)
Benzophenone	97	90
4-t-Butylcyclohexanone	92	80
Benzaldehyde	98	54
$\Delta^3$ -Cyclohexenecarboxaldehyde	96	79

#### 4. Selective reduction of phenolic rings

The last of the new methods to be discussed is a process which allows the reduction of phenolic rings under extremely mild conditions and which was expressly designed for application to the synthesis of carbocyclic natural products and other complex molecules. There is a need for such a procedure, since the presently known methods, high-pressure catalytic hydrogenation and alkali metal-ammonia (Birch) reduction, are subject to severe limitations in complex molecules with reducible or base-sensitive functional groups. The method is illustrated with *p*-cresol. In the first step the phenol is etherified to form the isobutyric acid (XXXIX) by reaction with chloretone

E. J. COREY

(XL) and base. Reaction of (XXXIX) as the sodium salt with bromine at  $-20^{\circ}$  in aqueous dimethyl sulfoxide produces a bromolactone (XL1) which can be isolated and converted by reaction with silver ion in methanol to a methoxy lactone (XLII). The methoxy diene lactone (XLII) is very easily reduced and is easily converted to a variety of derivatives of 4-methylcyclohexanone by standard mild procedures. This method has also been applied to polycyclic structures such as (XLIII).



*Many of the ideas presented here are generalizations of concepts and principles which have been developed and demonstrated by those who have been engaged in the synthesis of specific complex molecules over the past few decades. It is the hope of the author that this attempt to present the methods of organic synthesis in a general form will serve to emphasize the importance of their achievements as well as to clarify some of the underlying methodology.*

*The discussion in Part I has doubtless been influenced to some degree by the writer's personal experience. Other modes of thought, other interests might well have produced a different view, especially of the subsidiary details. It is hoped that a much more systematic, rigorous and complete account of the logic of Synthesis and the basic data with which it operates can be given in the near future.*

*Mention should be made to two superb general accounts of Synthesis, written by one of the Grand Masters<sup>34,35</sup>, which are highly recommended to those interested in a broad view of the subject.*

*Finally, the writer acknowledges his indebtedness to a splendid group of collaborators for the accomplishment of the work outlined in Part II. These gentlemen by name are: Dieter Seebach, Roland A. E. Winter, Francis A. Carey, Charles Cumbo, Gottfried Märkl, Tony Durst, George Kwiatkowski, Sandor Barcza and Georg Klotmann. The U.S. National Science Foundation and National Institutes of Health provided financial support.*

## References

- <sup>1</sup> F. Berlage, K. Bernauer, W. von Phillipsborn, P. Waser, H. Schmid, and P. Karrer. *Helv. Chim. Acta* **42**, 394 (1959).
- <sup>2</sup> E. J. Corey and R. R. Sauers. *J. Am. Chem. Soc.* **81**, 1739 (1959).

## METHODS FOR THE CONSTRUCTION OF COMPLEX MOLECULES

- <sup>3</sup> D. H. R. Barton, A. M. Deflorin, and O. E. Edwards. *J. Chem. Soc.* 530 (1956).
- <sup>4</sup> G. Stork, E. E. van Tamelen, L. J. Friedman, and A. Burgstahler. *J. Am. Chem. Soc.* **75**, 384 (1953).
- <sup>5</sup> R. B. Woodward, F. Sondheimer, D. Taub, K. Heusler, and W. M. MacLamore. *J. Am. Chem. Soc.* **74**, 4223 (1952).
- <sup>6</sup> W. S. Johnson, B. Bannister, and R. Pappo. *J. Am. Chem. Soc.* **78**, 6331 (1956).
- <sup>7</sup> J. B. Siddall, J. P. Marshall, A. Bowers, A. D. Cross, J. A. Edwards, and J. H. Fried. *J. Am. Chem. Soc.* **88**, 379, 862 (1966).
- <sup>8</sup> G. Stork, S. D. Darling, I. T. Harrison and P. S. Wharton. *J. Am. Chem. Soc.* **84**, 2018 (1962).
- <sup>9</sup> L. R. C. Barclay, C. E. Milligan, and N. D. Hall. *Can. J. Chem.* **40**, 1664 (1962).
- <sup>10</sup> P. von R. Schleyer, M. M. Donaldson, R. D. Nicholas and C. Cupas. *Org. Syntheses* **42**, 8 (1962).
- <sup>11</sup> E. J. Corey and W. R. Hertler. *J. Am. Chem. Soc.* **81**, 5209 (1959).
- <sup>12</sup> G. Cainelli, M. Lj. Mihailović, D. Arigoni, and O. Jeger. *Helv. Chim. Acta* **42**, 1126 (1959).
- <sup>13</sup> a F. D. Greene, M. L. Savitz, H. H. Lau, F. D. Osterholtz, and W. N. Smith. *J. Am. Chem. Soc.* **83**, 2196 (1961);  
b J. S. Mills and V. Petrow. *Chem. Ind. (London)* 946 (1961);  
c M. Akhtar and D. H. R. Barton. *J. Am. Chem. Soc.* **83**, 2213 (1961);  
d C. Walling and A. Padwa. *J. Am. Chem. Soc.* **83**, 2207 (1961).
- <sup>14</sup> D. H. R. Barton, J. M. Beaton, L. E. Geller, and M. M. Pechet. *J. Am. Chem. Soc.* **82**, 2640 (1960).
- <sup>15</sup> A. W. Burgstahler and I. C. Norden. *J. Am. Chem. Soc.* **83**, 198 (1961).
- <sup>16</sup> R. B. Woodward, F. E. Bader, H. Bickel, A. J. Frey, and R. W. Kierstead. *Tetrahedron* **2**, 1 (1958).
- <sup>17</sup> E. E. van Tamelen, L. J. Dolby, and R. G. Lawton. *Tetrahedron Letters* No. 19, 30 (1960).
- <sup>18</sup> D. H. R. Barton, G. W. Kirby, W. Steglich, and G. M. Thomas. *Proc. Chem. Soc.* 203 (1963).
- <sup>19</sup> F. H. Mosher and A. L. Thomas. *Phthalocyanine Complexes*, American Chemical Society Monograph 157, Reinhold Publishing Co., New York, N.Y.
- <sup>20</sup> D. H. Busch and M. C. Thompson. *J. Am. Chem. Soc.* **84**, 1762 (1962).
- <sup>21</sup> G. Wilke. *Angew. Chem. Int. Ed. Engl.* **2**, 105 (1963); **5**, 151 (1966).
- <sup>22</sup> A. Eschenmoser *et al.* *Angew. Chem. Int. Ed. Engl.* **3**, 490 (1964).
- <sup>23</sup> K. N. Menon and R. Robinson. *J. Chem. Soc.* 780 (1932).
- <sup>24</sup> A. R. Battersby. *Pure Appl. Chem.* **14**, 117 (1967).
- <sup>25</sup> E. J. Corey and S. Nozoe. *J. Am. Chem. Soc.* **87** 5733 (1965).
- <sup>26</sup> E. J. Corey, M. Ohno, R. B. Mitra, and P. A. Vatakencherry. *J. Am. Chem. Soc.* **86**, 478 (1964).
- <sup>27</sup> E. J. Corey and D. Seebach. *Angew. Chem. Int. Ed. Engl.* **4**, 1075, 1077 (1965).
- <sup>28</sup> E. J. Corey, D. Seebach, and B. W. Erickson. To be published.
- <sup>29</sup> E. J. Corey and R. A. E. Winter. *J. Am. Chem. Soc.* **85**, 2677 (1963).
- <sup>30</sup> E. J. Corey, F. A. Carey, and R. A. E. Winter. *J. Am. Chem. Soc.* **87**, 934 (1965).
- <sup>31</sup> See T. Mukaiyama and Y. Yokota. *Bull. Chem. Soc. Japan* **38**, 858 (1965).
- <sup>32</sup> E. J. Corey and G. Märkl. Abstracts, Nineteenth Organic Symposium, American Chemical Society, 1965, p. 7.
- <sup>33</sup> See R. L. Gay, S. Boatman, and C. R. Hauser. *Chem. Ind. (London)* 1789, (1965).
- <sup>34</sup> R. B. Woodward in *Perspectives in Organic Chemistry*, Ed. by A. R. Todd, Interscience Publishers, Inc., New York, N.Y., 1956, p. 155.
- <sup>35</sup> R. B. Woodward, *Pointers and Pathways in Research*, G. Hofsteizer for Ciba of India, Ltd., Bombay, India, 1963, p.23.

## **SECTION I: POLYSACCHARIDES**

### **INTRODUCTORY REMARKS BY THE HONORARY PRESIDENT OF THE SECTION**

J. K. N. JONES

The Scientific Programme Committee has asked me, as honorary section president, to present my views on polysaccharide chemistry, past and future. This I do with pleasure. The past I can vouch for—the future I can only guess at. Just 30 years ago E. L. Hirst accepted the Chair of Chemistry at Bristol University, U.K. I had just obtained the Ph.D. degree and was fortunate in that he invited me to come with him. At that time polysaccharide chemistry was in its infancy. The following figures show how this branch of chemistry has grown. In the American *Chemical Abstracts* Decennial Index (1937–1946), references to polysaccharide chemistry covered 2 pages, *i.e.*, about 480 references. References to starch covered five pages (1200 references), and those to cellulose and its derivatives, 20 pages (4800 references). A total of 6480 references, including patents, in a ten year period. During 1947–1956 references to polysaccharides covered 6½ pages, *i.e.* 1360 references, and the total number of references to cellulose and its derivatives, starch and polysaccharides was 10 640. During the period 1957–1961 (5 years) the references have doubled to about 9360 including 1680 on polysaccharides (many on muco-polysaccharides). This is an average of 2000 references per year, nearly six a day. The increase was due, in part, to the growing interest of biologists and biochemists in polysaccharide chemistry and to the growth of science in the U.S.S.R. and Japan. Who can hope to read all of these publications, even if only in abstract, and still get some bench work done? It will be noted that I have not included the nucleic acids, although they could also be considered as polysaccharide derivatives.

One of E. L. Hirst's interests was in the biological origin of the sugars. A hypothesis advanced was that pentoses resulted from hexuronic acids by decarboxylation. To test this hypothesis he decided to examine the structures of complex polymers which were built up of pentose, uronic acid and hexose residues. This was the beginning of a systematic chemical analysis of gums, mucilages and pectins. It soon became clear that the available methods, for a quantitative analysis of the methylated sugars resulting from hydrolysis of a methylated polymer, were inadequate.

J. K. N. JONES

When I first became interested in research I remember asking a graduate student what sort of research work was done in polysaccharide chemistry. The disgruntled reply was "All you do is methylate and hydrolyse". However, as is usually the case in times of depression and frustration, a new technique—paper chromatography—was developed and applied to the separation of carbohydrates, by Partridge. This put colour and glamour into carbohydrate chemistry and, more important, the purity of fractions of sugars and their derivatives could be assessed very quickly and on very small quantities. These are the advances I wish to emphasize—speed and the use of micro techniques; ever quicker and ever smaller. In 1936 a student might take 2–3 years to determine the chain length of starch using 50 g of polysaccharide. Now the operation can be carried out with greater accuracy in 3 days using 50 mg or less. Moreover, the present-day carbohydrate chemist knows that starch is a mixture, whereas this was a matter for dispute in 1936.

About 1939 we investigated the structure of certain plant gums by the methylation-fractional distillation procedure. A few years ago I had the opportunity of examining these fractions using paper chromatography and I was horrified to see how impure some of them were. I fear that some of the earlier work should be repeated, from a quantitative angle at least, and I am happy to see that this is now being done.

The next techniques used by polysaccharide chemists were periodate and lead tetraacetate oxidations, from which was developed the elegant Smith degradation procedure. These methods gave much new evidence, ambiguous in some cases and unsatisfactory in others. Nevertheless its use gave further information and enabled ideas on polysaccharide structure to be checked on a micro scale.

Enzymes were also employed to break down polymers and to isolate smaller fragments which were not obtainable by other methods. This procedure is open to some criticism, because of possible resynthesis, but it is perhaps superior to acetolysis techniques and partial hydrolysis by acids. The use of synthetic "enzymes" of the type used by T. J. Painter has now been developed.

It soon became clear that many polysaccharide preparations were mixtures and physical procedures were developed to facilitate the fractionation of these polymers. Paper electrophoresis, Tiselius type electrophoresis, ultra centrifugation, immuno-electrophoresis and other techniques were developed in order to test the homogeneity or, more usually, the lack of chemical and physical homogeneity of polysaccharide preparations.

Now gel electrophoresis, using both aqueous and non-aqueous solutions, gas liquid chromatography and mass spectrometry are used as means of separating and analysing polymers and their fragmentation products on an ever smaller scale. In 1948 W. H. Wadman built the first fraction collector to be used in conjunction with a cellulose column. This apparatus was made, in part, from a bomb release mechanism bought, *by weight*, from war surplus suppliers. Carbohydrate chemistry had become automated. Despite all these advances, with the possible exception of cellulose, it is doubtful whether the complete structure is yet known of any polysaccharide.

What of the future? The time is rapidly approaching when polysaccharide

## INTRODUCTORY REMARKS

analysis will be fully automated. Some 27 years ago Professor E. L. Hirst and I tried, unsuccessfully, to follow the fractional distillation of methylated sugars using an automatic collecting apparatus in which the refractive index could be followed without releasing the vacuum. It seems to me that it should now be possible to devise, *at a cost*, an automated apparatus which would check the purity of a polysaccharide and fractionate it if necessary. It would also analyse its sugar composition, periodate oxidize a portion of it, acetylate and/or methylate it and refractionate the methylated derivative. The determination of the physical constants of the fractions, such as the n.m.r. spectrum, viscosity, etc., and hydrolysis of each fraction could also be automated. The identities and quantities of the sugar derivatives, determined in each fraction by a finger print technique or by mass spectrometry, could all be recorded automatically at the end of the sequence. Dr. S. A. Barker and his colleagues have worked out other procedures in which one sugar at a time can be removed and identified and you will hear more about this later.

The finger print technique involves the preparation of every one of the possible methyl ethers of all the known common sugars. These are examined under standard conditions, by paper chromatography, as their reducing sugars, or by thin layer chromatography and by gas-liquid chromatography as their methyl glycoside acetates, as the glycitol acetate derivatives and as the lactone acetate or tri- or dimethyl silyl derivatives. Each compound is expected to have a characteristic pattern of colour and rates of movement and the results would be scanned, stored and analyzed in the computer. So far D-glucose, L-arabinose, and D-xylose derivatives have been prepared and could be utilized in this fashion. To build such an apparatus and to pay for technical assistance would be very expensive but I believe that the knowledge is now available to make it. Technologies which allow the landing of television transmitters on the moon should have no difficulty in solving this problem.

Great progress has been made in the determination of the structure of polysaccharides. The biological origin and function of these polymers, however, has only attracted attention, mainly from biologists, within the last few years. Little is known of the biosynthesis of complex polymers such as gums and mucilages and it is likely that big advances will be made in this field and in the chemical synthesis of polysaccharides within the next few years.

Our speaker this afternoon, Dr. G. O. Aspinall knows as much about the structure and chemistry of gums and mucilages as anyone. He started as an aromatic chemist but since 1948 has been engaged in studying the chemistry of carbohydrates. He has done a lot of work on the use of the gas-liquid chromatographic apparatus in the analysis of methylated sugars obtained from gums and mucilages. He is known internationally for his work on polysaccharides and has given lectures, by invitation, in East and West Europe and at centres in North America. Today he will indicate the advances that have been made, mainly by his colleagues and himself, in this branch of polysaccharide chemistry, at the University of Edinburgh, and he will show that some order can be developed from apparent chaos.

# THE EXUDATE GUMS AND THEIR STRUCTURAL RELATIONSHIP TO OTHER GROUPS OF PLANT POLYSACCHARIDES

G. O. ASPINALL

*Department of Chemistry, The University, Edinburgh 9, Scotland*

The exudate gums are complex acidic polysaccharides produced by trees either spontaneously or after mechanical injury. Each polysaccharide contains three or more constituent sugars including D-glucuronic acid (or its 4-methyl ether) and/or D-galacturonic acid. Highly branched structures are invariably encountered and since each constituent sugar may be present in furanose or pyranose ring forms may have different configurations at its glycosidic linkages, and may be involved in a variety of types of linkage, the elucidation of the detailed molecular structure of the exudate gums provides one of the most challenging problems in organic chemistry.

Many plant gums consist almost entirely of polysaccharides and consequently the problem of isolating these substances free from other materials, especially from other natural macromolecules, is often relatively simple compared with that of the isolation of other plant polysaccharides which are present as cell-wall components. The additional problem of obtaining polysaccharides in a sufficiently homogeneous form for detailed chemical examination cannot be considered at length on this occasion. Suffice it to say that three types of polymer heterogeneity have been recognized amongst the polysaccharide constituents of the exudate gums. First, there are those gums, such as gum tragacanth<sup>1</sup> and *Khaya senegalensis* gum<sup>2,3</sup>, in which two polysaccharide components of entirely different structural type are present. Secondly, at the other end of the scale, there is the kind of micro-heterogeneity found in *Combretum leonense* gum<sup>4</sup> in which polysaccharide sub-fractions show small differences in composition but no differences in the nature of the structural units or in the linkages between them. Thirdly, heterogeneity of an intermediate type has been encountered recently in *Anogeissus leiocarpus* (formerly *A. schimperi*) gum<sup>5,6</sup>. This gum contains two discrete polysaccharide components which are sufficiently different to permit fractionation on a preparative scale. Structural investigations so far carried out indicate that the two polysaccharides contain many of the same structural units but in markedly different relative proportions.

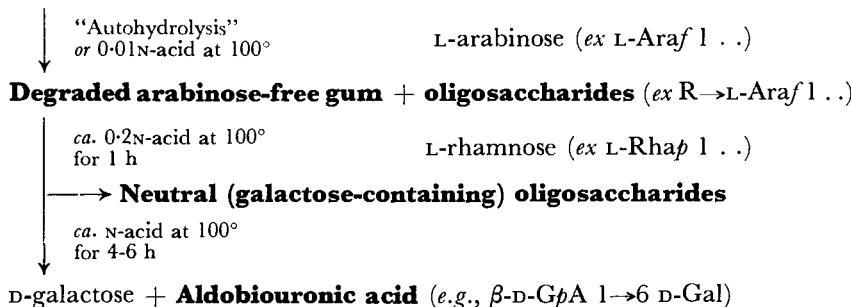
To an increasing extent complex plant polysaccharides may be classified in terms of families of related molecular species in which the members of each group contain common structural units, most frequently the sequences of sugar residues in the interior chains, but differ considerably in the nature and number of other units attached as side-chains. This type of structural classification of polysaccharides, often of diverse origins, has been applied to many of the cell-wall polysaccharides of higher plants<sup>7,8</sup>, the xylan group

## G. O. ASPINALL

providing the most fully authenticated example. I wish today to extend this type of classification to three groups of polysaccharides amongst the exudate gums of which sufficient is known to permit reasonably detailed structural formulations. In the case of two of these groups this kind of systematization serves to emphasize the structural relationship between certain of the exudate gums and other types of plant polysaccharides.

The presence in many gum polysaccharides of a variety of glycosidic linkage which undergo cleavage at markedly different rates has permitted the use of controlled acid hydrolysis for relatively selective fragmentation. This type of graded hydrolysis, which was first used over twenty-five years ago by the late F. Smith in studies on gum arabic, and by E. L. Hirst and J. K. N. Jones in studies on damson and cherry gums<sup>9</sup>, still forms the basis for the controlled fragmentation of gums. The power of this method of approach has been greatly extended during more recent years by modern chromatographic methods for the separation of oligosaccharides, and the general procedure is outlined in *Figure 1*. More recently, alternative methods of fragmentation have been exploited; these include acetolysis, which often results in markedly different "cracking patterns" from those of partial hydrolysis in aqueous solution, and the Smith degradation of sequential periodate oxidation, borohydride reduction, and mild acid hydrolysis of resulting acyclic acetals, which provides a powerful method for the isolation of those portions of the molecule which are resistant to oxidative glycol cleavage. Advantage has also been taken of changes in the relative rates of cleavage of different glycoside bonds consequent on structural modifications in polysaccharides such as reduction of or oxidation to hexuronic acid residues. In all phases of the examination of complex macromolecules, from the polysaccharides themselves to oligosaccharides as degradation products, the classical methylation procedure has continued to be widely exploited for the determination of sites of substitution, the more especially since gas chromatographic methods have been developed for the identification of mixtures of methyl glycosides of methylated and partially methylated sugars.

### Gum acid

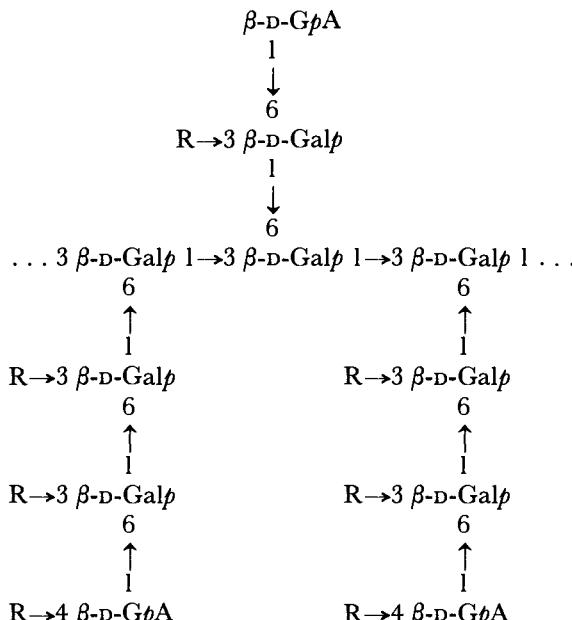


*Figure 1.* Graded acid hydrolysis of gums

We may consider first the galactan group of exudate gums in which each polysaccharide contains a branched core of  $\beta$ -D-galactopyranose residues

## EXUDATE GUMS AND OTHER PLANT POLYSACCHARIDES

mutually joined by  $1 \rightarrow 3'$  and  $1 \rightarrow 6'$  linkages. The distribution of the two types of linkage is such that those of the  $1 \rightarrow 3'$  type predominate in the interior chains whilst those of the  $1 \rightarrow 6'$  type are concentrated in the outer chains. Gum arabic, one of the best known exudate gums, produced by trees of *Acacia senegal* (Leguminosae, Mimosaceae) in the Sudan, although one of the most complex known polysaccharides, was the first gum of this type in which the nature of the galactan core was established. The main structural features of the gum, which are based on the application of the general methods previously outlined, are summarized in *Figure 2*. It must be emphasized, however, that this highly branched type of structure does not provide a unique representation of this complex polysaccharide, and, if anything, is an over-simplification.



*Figure 2.* Gum arabic, where  $R = L\text{-Araf}$   $1 \dots$ ,  $L\text{-Rhap}$   $1 \dots$ ,  $\alpha\text{-D-Galp}$   $1 \rightarrow 3 L\text{-Araf}$   $1 \dots$ , or occasionally  $\beta\text{-L-Arap}$   $1 \rightarrow 3 L\text{-Araf}$   $1 \dots$

Of the various acid-labile units ( $R$ ) in the outer chains of gum arabic only the *L*-rhamnopyranose residues have been placed with certainty. Although *L*-rhamnopyranosidic linkages are readily hydrolysed in aqueous solution, they are sufficiently resistant to cleavage by acetolysis to permit the isolation by this procedure of oligosaccharides with intact *L*-rhamnosidic linkages. Thus, acetolysis of carboxyl-reduced arabic acid, followed by deacetylation of the products, led to the isolation of  $4\text{-O-}\alpha\text{-L-rhamnopyranosyl-D-glucose}$  and  $O\text{-}\alpha\text{-L-rhamnopyranosyl-(1}\rightarrow 4\text{)-O-}\beta\text{-D-glucopyranosyl-(1}\rightarrow 6\text{)-D-galactose}$  (see *Figure 3*)<sup>10</sup>.

Gums containing the same type of basal structure have been recognized from other *Acacia* species and from a number of botanically unrelated trees. These polysaccharides vary to some extent in the degree of branching in the galactan core and also differ in the sites of attachment of the peripheral

## G. O. ASPINALL

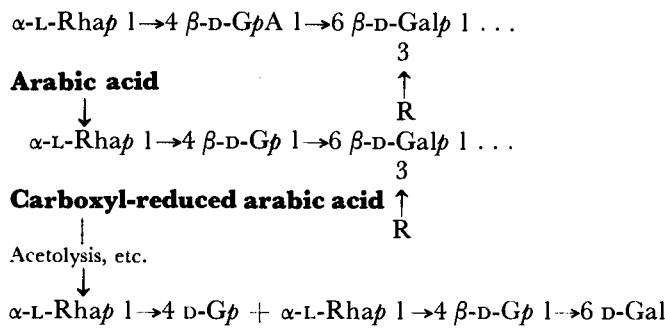


Figure 3. Location of L-rhamnopyranose end groups in arabic acid

L-arabinofuranose and L-rhamnopyranose residues and in the nature of the more complex substituted L-arabinofuranose units. Polysaccharides containing a highly branched galactan core are also found in the wood of many coniferous trees, especially of larches. These polysaccharides generally possess somewhat less highly ramified structures and contain a lower proportion of sugar residues other than D-galactose, than do the exudate gums. On structural grounds, however, there is no clear line of demarcation between the two groups of polysaccharides. *Table 1* indicates the similarities

*Table 1.* Sugar residues in the outer chains of some gums and wood polysaccharides containing a branched galactan core

<i>Source</i>	<i>β-D-Glucuronic acid residues linked 1→6' to D-galactose residues in outer chains</i>	<i>Peripheral units readily cleaved on mild hydrolysis</i>
<b>GUMS</b>		
<i>Acacia senegal</i> <sup>9,10</sup> (gum arabic)	$\alpha\text{-L-Rhap} 1 \rightarrow 4 \beta\text{-D-GpA} 1 \dots$ $\beta\text{-D-GpA} 1 \dots$	$\text{L-Rhap} 1.$ $\text{L-Araf} 1.$ $\alpha\text{-D-Galp} 1 \rightarrow 3 \text{ L-Araf} 1.$ $\beta\text{-D-Arap} 1 \rightarrow 3 \text{ L-Araf} 1.$
<i>Acacia pycnantha<sup>11</sup></i>	$\beta\text{-D-GpA} 1 \dots$	$\text{L-Rhap} 1.$ $\text{L-Araf} 1.$ $\text{L-Araf} 1 \rightarrow 3 \text{ L-Araf} 1.$
<i>Ferula</i> spp. <sup>12</sup> (asafotida gum)	$\beta\text{-D-GpA} 1 \dots$ 4-Me $\beta\text{-D-GpA} 1 \dots$	$\text{L-Araf} 1.$
<i>Araucaria bidwillii</i> <sup>13</sup>	$\beta\text{-D-GpA} 1 \dots$ 4-Me $\beta\text{-D-GpA} 1 \dots$	$\text{L-Rhap} 1.$ $\text{L-Araf} 1.$

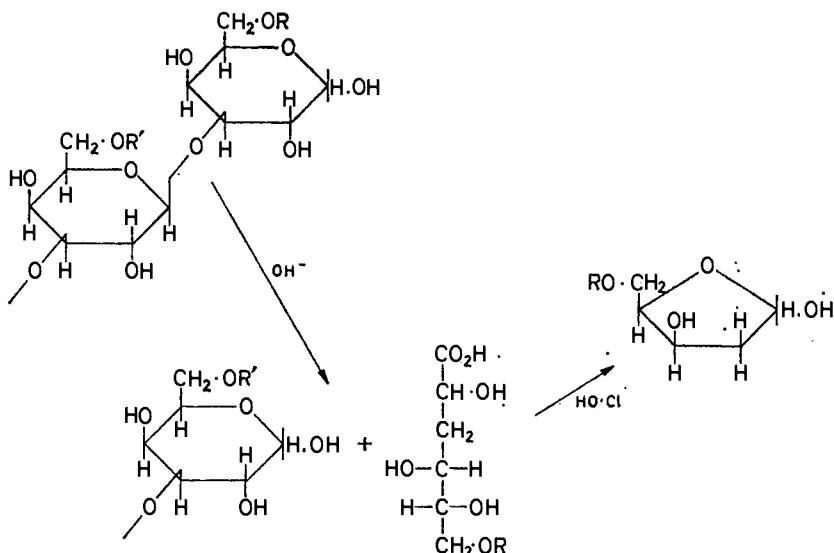
## WOOD POLYSACCHARIDES

European, Western, and Japanese larches <sup>14</sup> ( <i>Larix</i> <i>decidua</i> , <i>L. occidentalis</i> , <i>L. leptolepis</i> )		$\text{L-Araf} 1.$ $\beta\text{-L-Arap} 1 \rightarrow 3 \text{ L-Araf} 1.$
Mountain larch <sup>15</sup> ( <i>L. lyallii</i> )	$\beta\text{-D-GpA} 1 \dots$	$\text{L-Araf} 1.$ $\beta\text{-L-Arap} 1 \rightarrow 3 \text{ L-Araf} 1.$
Maritime pine <sup>16</sup> ( <i>Pinus pinaster</i> )	$\beta\text{-D-GpA} 1 \dots$	$\text{L-Araf} 1.$ $\beta\text{-L-Arap} 1 \rightarrow 3 \text{ L-Araf} 1.$ $\alpha\text{-D-Xylp} 1 \rightarrow 3 \text{ L-Araf} 1.$

## EXUDATE GUMS AND OTHER PLANT POLYSACCHARIDES

in the nature (although not necessarily in the sites of attachment) of the sugar residues in the outer chains of some of the exudate gums and coniferous wood arabinogalactans.

One of the major problems awaiting solution in structural studies on polysaccharides of the galactan group is that of the unambiguous location of the various acid-labile peripheral units. One possible method of attack, presently under investigation, takes advantage of the susceptibility of 3-O-substituted sugars to degradation by alkali to form metasaccharinic acids. Since these polysaccharides contain interior chains of 3-O-substituted D-galactopyranose residues which carry side-chains attached at C-6, alkaline degradation should furnish a series of O-glycosyl saccharinic acids. In order to avoid unwanted complications arising from the formation of stereoisomeric saccharinic acids and to minimize the possibility of inadvertent autohydrolysis these acids have been degraded by treatment with hypochlorite at pH 5 to the corresponding 5-O-substituted 2-deoxy-D-threo-pentoses, and the separation and characterization of these oligosaccharides is under investigation. The degradative scheme is outlined in *Figure 4*<sup>17</sup>.

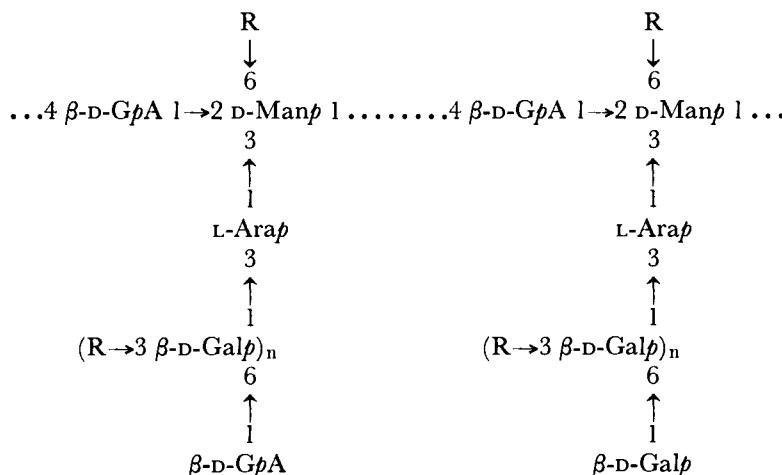


*Figure 4.* Alkaline degradation of 1→3' linked galactans

The second group of exudate gums to which I wish to direct attention may be classified, in terms of the inner chains to which other units are attached as side-chains, as glucuronomannans. These gums, like those of the galactan group, usually contain acid-labile pentose residues on the periphery of the molecule and may be subjected to the same type of controlled hydrolysis to give degraded polysaccharides and various mixtures of neutral and acidic oligosaccharides. Gum ghatti of Indian origin from *Anogeissus latifolia* (Combretaceae) provided the first example of a polysaccharide in which clear evidence was obtained for the presence of interior chains of D-glucuronic acid and D-mannose residues, and some of the main structural

G. O. ASPINALL

features of the gum are shown in *Figure 5*<sup>18,19</sup>. Since partial acid hydrolysis of the gum had yielded two polymer-homologous series of 1→6' linked D-galactose-containing oligosaccharides, the one series with and the other without a 3-O-substituted L-arabinose reducing group, our first inclinations were to suppose that these units arose from main chains of  $\beta$ -D-galactopyranose residues interrupted at intervals by L-arabinose residues, and that units of the two aldobiouronic acids, 2-O-( $\beta$ -D-glucopyranosyluronic acid)-D-mannose and 6-O-( $\beta$ -D-glucopyranosyluronic acid)-D-galactose, were attached thereto as side-chains. Evidence for the type of structure shown in *Figure 5* came from fragmentation of the gum involving two successive degradations by Smith's procedure followed by partial acid hydrolysis which afforded a mixture of oligosaccharides including 6-O- $\beta$ -D-galactopyranosyl-D-galactose, 3-O- $\beta$ -D-galactopyranosyl-L-arabinose, and 3-O-L-arabinopyranosyl-D-mannose<sup>20</sup>. The proposed structure for gum ghatti is one in which these disaccharides have their origin in side-chains of 6-O-substituted  $\beta$ -D-galactopyranose residues linked via 3-O-substituted L-arabinopyranose to C-3 of D-mannopyranose residues in the interior structure of the gum. At present there are only indirect indications that some of these side-chains are terminated by D-glucuronic acid end groups, and thus provide the source of one of the two aldobiouronic acids, 6-O-( $\beta$ -D-glucopyranosyluronic acid)-D-galactose, formed on partial hydrolysis. Evidence that the second aldobiouronic acid, 2-O-( $\beta$ -D-glucopyranosyluronic acid)-D-mannose, originates from blocks of D-glucuronic acid and D-mannose residues rather than from isolated disaccharide units within the structure has been obtained recently by the isolation on partial hydrolysis of a tetrasaccharide containing alternating residues of the two sugars<sup>21</sup>.

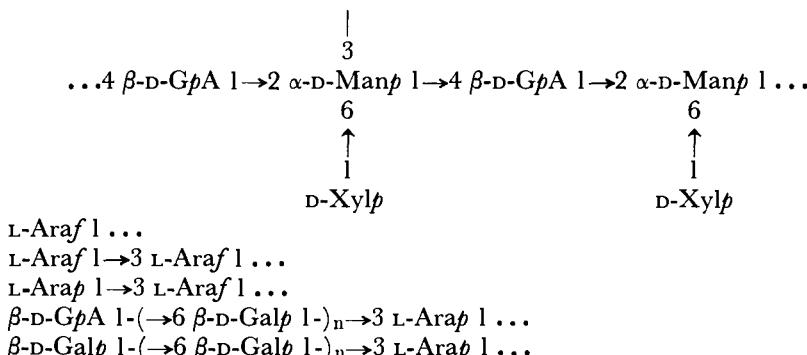


*Figure 5.* Gum ghatti, where the majority of sites indicated carry substituents R = L-Araf 1 . . . , or less frequently, L-Araf 1 → 2 L-Araf 1 . . . , L-Araf 1 → 3 L-Araf 1 . . . or L-Araf 1 → 5 L-Araf 1 . . .

More convincing evidence still for the presence in a polysaccharide of main chains composed of residues of D-glucuronic acid and D-mannose has been provided in recent studies on the gum from a botanically related tree,

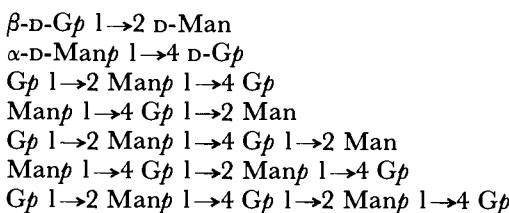
## EXUDATE GUMS AND OTHER PLANT POLYSACCHARIDES

*Anogeissus leiocarpus* (formerly *A. schimperi*), of West African origin. Our first investigations of this gum showed that it gave rise on partial hydrolysis to largely the same neutral and acidic oligosaccharides as gum ghatti, but that the relative amounts of the various oligosaccharides formed were substantially different in the two cases<sup>5</sup>. Further examination of this gum has shown that it contains two distinct but structurally related polysaccharide components<sup>6</sup>. The major component, leiocarpan A, which is selectively precipitated with cetyltrimethylammonium bromide, contains D-glucuronic acid, D-mannose, L-arabinose, and D-xylose as the main constituent sugars and a very low proportion (*ca.* 4 per cent) of D-galactose residues. The main structural features of the polysaccharide are outlined in *Figure 6*. The pentoses are largely present as single unit L-arabinofuranose and D-xylopyranose side-chains, and the main chains of D-glucuronic acid and D-mannose residues are much more readily exposed than in gum ghatti where these chains carry long side-chains and are hidden in the interior of the molecular structure. The D-galactose residues, although present in only low proportion, are probably present in the same type of side-chains as in gum ghatti since they give rise to the same oligosaccharides on partial hydrolysis. These side-chains in leiocarpan A, however, are probably shorter than in gum ghatti and attached much less frequently.



*Figure 6.* Partial structures for leiocarpan A

Evidence for the nature of the interior chains in leiocarpan A and proof that residues of 4-*O*-substituted  $\beta$ -D-glucuronic acid and 2-*O*-substituted  $\alpha$ -D-mannopyranose are present largely, if not exclusively, in a regularly alternating sequence has been obtained by the isolation from acetolysis of the carboxyl-reduced polysaccharide of the series of neutral oligosaccharides shown in *Figure 7*<sup>22</sup>.



*Figure 7.* Oligosaccharides from carboxyl-reduced leiocarpan A.

G. O. ASPINALL

Several plant gums, including some from the *Prunus* genus (Rosaceae), give rise to the aldobiouronic acid, 2-O-( $\beta$ -D-glucopyranosyluronic acid)-D-mannose, on partial hydrolysis, but it is not yet clear whether these gums can be properly regarded as members of the glucuronomannan family of polysaccharides. It is significant, however, that Zitko and his collaborators<sup>23</sup> in Bratislava have isolated a series of oligosaccharides composed of glucuronic acid and mannose units from the partial hydrolysis of apricot gum pointing to the presence of blocks of these units in some part of the structure of the gum. The glucuronomannan type of structure is at present unknown amongst plant polysaccharides other than the exudate gums.

We may turn finally to an entirely different group of exudate gums, named by reference to their interior or basal chains as galacturonans or galacturonorhamnans. It is perhaps more correct to consider these gums as members of three sub-groups between which there is partial structural overlap. These gums are noteworthy in containing residues of both D-galacturonic acid and D-glucuronic acid (or its 4-methyl ether). The former sugar units are located mainly in the interior chains, but the latter are found as terminal units in side-chains attached to a variety of different sugar residues. The first sub-group of gums of the galacturonorhamnan family is represented by gums of the *Sterculia* genus (Sterculiaceae) of widely different origins in India (*S. urens*), West Africa (*S. setigera*) and Australia (*S. caudata*, syn. *Brachychiton diversifolium*), and by the gum from the botanically unrelated *Cochlospermum gossypium* (Bixineae)<sup>24-26</sup>. These gums occur naturally as partially acetylated polysaccharides and partial structures for the essentially similar parent polysaccharides are shown in Figure 8. Direct evidence for the presence of D-glucuronic acid residues as single unit side-chains follows from methylation studies and from the characterization of O-( $\beta$ -D-glucopyranosyluronic acid)-(1 $\rightarrow$ 3)-O-( $\alpha$ -D-galactopyranosyluronic acid)-(1 $\rightarrow$ 2)-L-rhamnose as one of the products of partial acid hydrolysis.

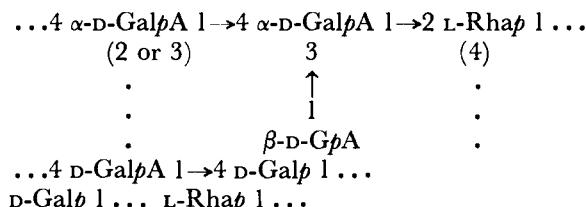
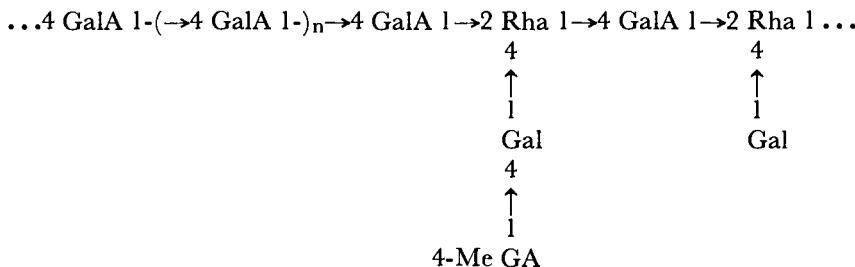


Figure 8. Partial structures for gums from *Sterculia* spp. and from *Cochlospermum gossypium*

The second sub-group of exudate gums of the galacturonorhamnan family is represented by those from the *Khaya* genus (Meliaceae)<sup>2,27,28</sup>. Like the *Sterculia* gums the *Khaya* gums occur naturally as partly acetylated polysaccharides, and the major polysaccharide components contain many of the same structural units, although they differ in detailed molecular architecture. The interior chains possess substantial blocks of contiguous D-galacturonic acid residues between which are interposed L-rhamnose residues. As in the *Sterculia* gums, the residues of D-glucuronic acid (as the 4-methyl ether) are present as end groups, but in the *Khaya* gums these units are joined to

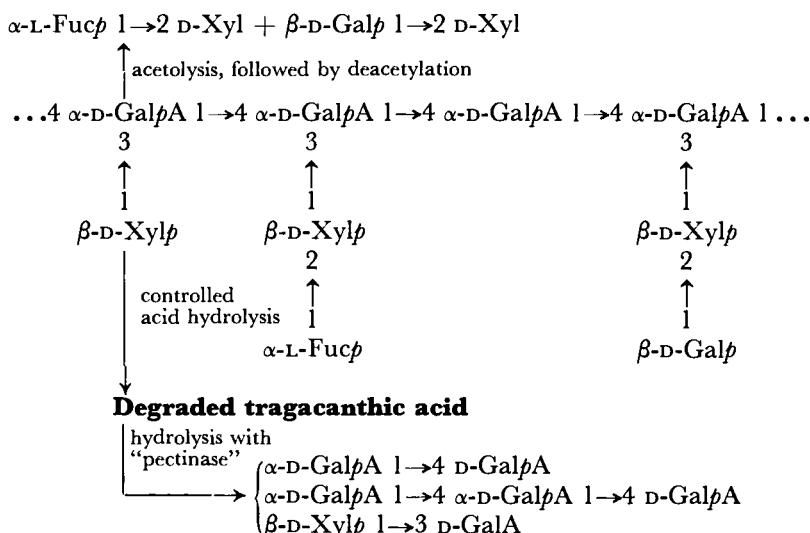
## EXUDATE GUMS AND OTHER PLANT POLYSACCHARIDES

D-galactose residues, which are probably linked in turn to L-rhamnose branching points (*Figure 9*).



*Figure 9.* Major polysaccharide components of gums from *Khaya* spp.

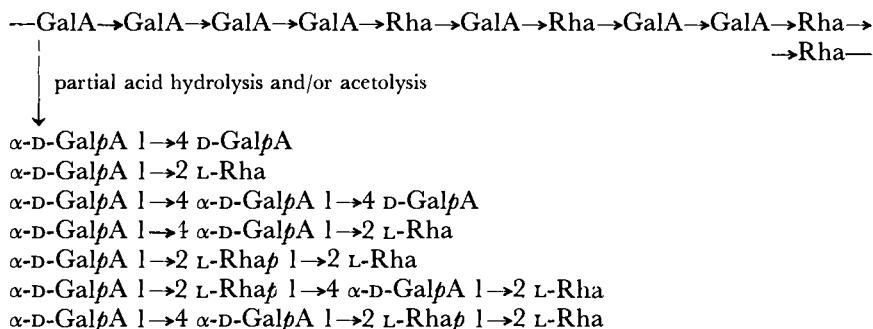
Tragacanthic acid, the main polysaccharide component of gum tragacanth from *Astragalus gummifer* (Leguminosae), is the only known representative of the third sub-group. In this polysaccharide the main chains are composed mainly of 1→4' linked α-D-galacturonic acid residues, although the recent isolation of small amounts of the aldobiouronate, 2-O-(α-D-galactopyranosyluronic acid)-L-rhamnose, as a partial hydrolysis product<sup>29</sup> suggests that the main chains may be interrupted occasionally by L-rhamnose residues. The short side-chains consist of either single β-D-xylopyranose or substituted 2-O-α-L-fucopyranosyl-D-xylopyranose or 2-O-β-D-galactopyranosyl-D-xylopyranose units. The key degradations leading to the assignment of the main structural features of the polysaccharide are shown in *Figure 10*<sup>1</sup>. It is noteworthy that L-fucopyranosidic, like L-rhamnopyranosidic, linkages are readily hydrolysed in aqueous solution, but are sufficiently stable to acetolysis to permit the isolation of 2-O-α-L-fucopyranosyl-D-xylose as one of the products. D-Glucuronic acid is only a trace constituent of tragacanthic acid and probably terminates some of the side-chains<sup>29</sup>.



*Figure 10.* Degradations of tragacanthic acid

G. O. ASPINALL

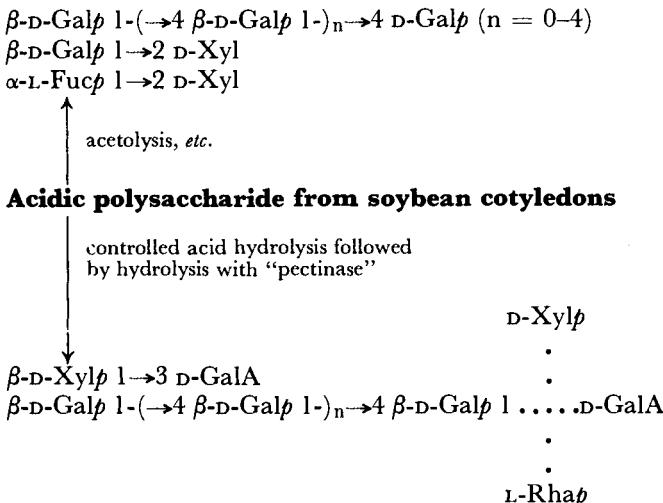
The structural relationship of tragacanthic acid to the pectins is immediately apparent in the interior galacturonan chains. Gums of the other two sub-groups, however, are related to pectins not only in containing similarly linked blocks of D-galacturonic acid residues but also in that these blocks are interposed by L-rhamnose residues. Galacturonans, as strictly homopolysaccharides, are now recognized to be of relatively infrequent occurrence, and several pectins, e.g., those from lemon peel<sup>30</sup>, lucerne (alfalfa)<sup>31</sup> and the bark of *Amabilis* fir<sup>32</sup> give rise on partial hydrolysis to 2-O-( $\alpha$ -D-galactopyranosyluronic acid)-L-rhamnose and higher oligosaccharides composed of these sugars. It now appears probable that the galacturonan chains in most pectins are interrupted at intervals by L-rhamnose residues. Recent investigations on the pectin-like acidic polysaccharides from soybean cotyledon meal<sup>33</sup> and soybean hulls<sup>34</sup> have provided most striking examples of polysaccharides in which different areas of the galacturonorhamnan chains may contain (a) blocks of galacturonic acid residues, (b) sequences of at least two rhamnose residues, and (c) alternating galacturonic acid and rhamnose units. The acidic oligosaccharides isolated from the soybean polysaccharides by partial hydrolysis or by acetolysis and the types of sequence from which they must arise are shown in *Figure 11*.



*Figure 11.* Fragmentation of the interior chains of soybean polysaccharides

The structural relationship between the soybean polysaccharides and the galacturonan-galacturonorhamnan group of exudate gum is not, however, limited to the interior portions of the molecular structure. The outer chains include the same single  $\beta$ -D-xylopyranose units linked 1- $\rightarrow$ 3' to D-galacturonic acid, and 2-O- $\alpha$ -L-fucopyranosyl-D-xylopyranose and 2-O- $\beta$ -D-galactopyranosyl-D-xylopyranose units as tragacanthic acid. To complicate the structural picture yet further the soybean polysaccharides contain linear chains of  $\beta$ -D-galactopyranose residues linked 1- $\rightarrow$ 4' as in the pectic galactan from *Lupinus albus* seeds<sup>35</sup>. These chains, however, are integral parts of an acidic polysaccharide rather than of a contaminating neutral galactan since stepwise degradation furnishes acidic oligosaccharides (not yet fully characterized) in which these chains are linked either directly or through xylose or rhamnose residues to a single galacturonic acid reducing unit. Fragmentations of the side-chains in the soybean cotyledon polysaccharide are summarized in *Figure 12*.

## EXUDATE GUMS AND OTHER PLANT POLYSACCHARIDES



*Figure 12.* Fragmentations of the side-chains of soybean cotyledon polysaccharide

In terms of their basal chains as galacturonorhamnans in which the proportions of galacturonic acid to rhamnose residues vary from *ca.* 3:1 in the *Sterculia* and *Khaya* gums to possibly 100:1 in tragacanthic acid, the three sub-groups comprise a broad spectrum of chemically related polysaccharides. The structural inter-relationships with each other and with the pectins are summarized in *Figure 13*.

The search for a complete structural basis for the classification of the highly branched polysaccharides from the exudate gums is still in its early stages. The three groups of polysaccharides to which I have directed attention may need to be extended or modified to accommodate the discovery of polysaccharides with further variations in detailed structure, and it is probable that entirely new structural groups of exudate gums will be recognized. There are indeed several well-known exudate gums, *e.g.*, the various *Prunus* gums<sup>9</sup>, mesquite gum<sup>9</sup>, and cholla gum<sup>9,36</sup>, which have not been considered in this survey. Although much is known of certain aspects of the chemistry of these gums, the evidence in my opinion, is not yet sufficient to provide a clear indication of which sequences of sugar units comprise the basal chains of the polysaccharides and of which are attached as side-chains to these interior chains.

The exudate gums are end products of plant metabolism, but in two of the groups to which I have referred the structural relationships to other types of plant polysaccharides indicate that similar biosynthetic pathways are followed in their formation. It is indeed possible in some cases that in gum formation polysaccharides which are already present in the plant, *e.g.*, as cell-wall constituents, undergo the apposition of additional sugar residues to the outer chains to give the yet more complex polysaccharides which are exuded to protect the injured plant against desiccation in hot dry climates and against infection by micro-organisms. Speculation aside, it is clear that advances in knowledge of the nature of plant heteropolysaccharides will come from parallel studies on the exudate gums as end products and on the

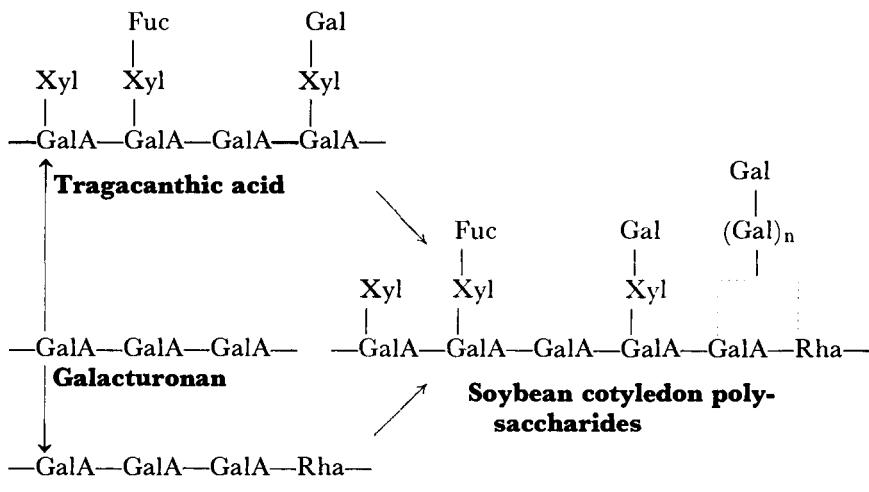
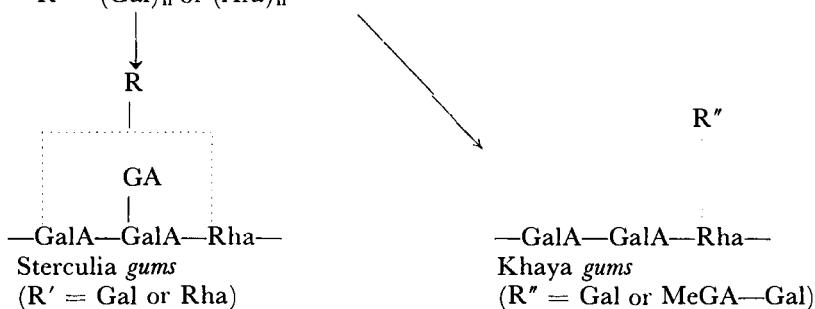
**Complex pectins** (e.g., citrus) $R = (Gal)_n$  or  $(Ara)_n$ 

Figure 13. Structural inter-relationships between polysaccharides of the galacturonan-galacturonorhamnan group

various cell-wall polysaccharides which are involved in the normal metabolism of the plant.

*In conclusion, it is a pleasure to acknowledge my personal gratitude to Professor Sir Edmund Hirst for his active interest and encouragement in all phases of the work carried out in Edinburgh, and to my various collaborators who have performed skilled experimental work and have also contributed in many fruitful discussions to the formulation of the conclusions outlined in the present review. Special mention should be made of the following people from whose unpublished results I have freely quoted: Drs. R. Begbie, Sylvia V. Egan, R. N. Fraser, J. M. McNab, I. M. Morrison, J. L. Whyte, J. N. C. Whyte, and Messrs. A. K. Bhattacharjee, J. J. Carlyle, I. W. Cottrell, J. W. T. Craig, D. B. Davies, J. P. McKenna, J. A. Molloy, G. R. Sanderson, M. Uddin, and C. C. Whitehead.*

## EXUDATE GUMS AND OTHER PLANT POLYSACCHARIDES

### References

- <sup>1</sup> G. O. Aspinall and J. Baillie. *J. Chem. Soc.* 1702 (1963).
- <sup>2</sup> G. O. Aspinall, M. J. Johnston, and A. M. Stephen. *J. Chem. Soc.* 4918 (1960).
- <sup>3</sup> G. O. Aspinall, M. J. Johnston, and R. Young. *J. Chem. Soc.* 2701 (1965).
- <sup>4</sup> G. O. Aspinall and V. P. Bhavanandan. *J. Chem. Soc.* 2685, 2693 (1965).
- <sup>5</sup> G. O. Aspinall and T. B. Christensen. *J. Chem. Soc.* 3461 (1961).
- <sup>6</sup> G. O. Aspinall, J. J. Carlyle, J. M. McNab, and A. Rudowski. Unpublished results.
- <sup>7</sup> G. O. Aspinall. *Chim. Biochim. Lignine Cellulose Hemicelluloses, Actes Symp. Intern.*, p. 421, Les Imprimeries Reunies de Chambéry, France (1964).
- <sup>8</sup> T. E. Timell. *Advances in Carbohydrate Chemistry*, **19**, 247 (1964); **20**, 409 (1965).
- <sup>9</sup> For a comprehensive survey of the chemistry of gums until 1958 approximately see F. Smith and R. Montgomery. *The Chemistry of Plant Gums and Mucilages*. Reinhold Publishing Corp., New York (1959).
- <sup>10</sup> G. O. Aspinall, A. J. Charlson, E. L. Hirst, and R. Young. *J. Chem. Soc.* 1696 (1963); G. O. Aspinall and R. Young. *J. Chem. Soc.* 3003 (1965).
- <sup>11</sup> G. O. Aspinall, E. L. Hirst, and A. Nicolson. *J. Chem. Soc.* 1697 (1959).
- <sup>12</sup> J. K. N. Jones and G. H. S. Thomas. *Canad. J. Chem.* **39**, 192 (1961).
- <sup>13</sup> G. O. Aspinall and R. M. Fairweather. *Carbohydrate Research*, **1**, 83 (1965).
- <sup>14</sup> For references see G. O. Aspinall. *Chim. Biochim. Lignine, Cellulose Hemicelluloses, Actes Symp. Intern.*, p. 89, Les Imprimeries Reunies de Chambéry, France (1964). See also Ref. 8.
- <sup>15</sup> J. K. N. Jones and P. E. Reid. *J. Polymer Sci.* Part C, No. 2, 63 (1963).
- <sup>16</sup> A. J. Roudier and L. Eberhard. *Bull. Soc. Chim.* 844 (1963).
- <sup>17</sup> G. O. Aspinall and J. P. McKenna. Unpublished results.
- <sup>18</sup> G. O. Aspinall, E. L. Hirst, and A. Wickstrøm. *J. Chem. Soc.* 1160 (1955).
- <sup>19</sup> G. O. Aspinall, B. J. Auret, and E. L. Hirst. *J. Chem. Soc.* 221, 4408 (1958).
- <sup>20</sup> G. O. Aspinall, V. P. Bhavanandan, and T. B. Christensen. *J. Chem. Soc.* 2677 (1965).
- <sup>21</sup> G. O. Aspinall, J. J. Carlyle, and C. C. Whitehead. Unpublished results.
- <sup>22</sup> G. O. Aspinall and J. M. McNab. *Chem. Commun.* 565 (1965).
- <sup>23</sup> V. Zitko, J. Rosik, M. Brutenicova, and J. Kubala. *Coll. Czech Chem. Comm.* **30**, 3501 (1965).
- <sup>24</sup> G. O. Aspinall and Nasir-ud-din. *J. Chem. Soc.* 2710 (1965).
- <sup>25</sup> G. O. Aspinall and R. N. Fraser. *J. Chem. Soc.* 4318 (1965).
- <sup>26</sup> G. O. Aspinall, R. N. Fraser, and G. R. Sanderson. *J. Chem. Soc.* 4325 (1965).
- <sup>27</sup> G. O. Aspinall, E. L. Hirst, and N. K. Matheson. *J. Chem. Soc.* 989 (1956).
- <sup>28</sup> G. O. Aspinall and A. K. Bhattacharjee. Unpublished results.
- <sup>29</sup> G. O. Aspinall, D. B. Davies, and R. N. Fraser. Unpublished results.
- <sup>30</sup> G. O. Aspinall, J. W. T. Craig, and J. L. Whyte. Unpublished results.
- <sup>31</sup> G. O. Aspinall and R. S. Fanshawe. *J. Chem. Soc.* 4215 (1961); G. O. Aspinall, J. A. Molloy, and M. Uddin. Unpublished results.
- <sup>32</sup> T. E. Timell and S. S. Bhattacharjee. *Canad. J. Chem.* **43**, 758 (1965).
- <sup>33</sup> G. O. Aspinall, R. Begbie, I. W. Cottrell, S. V. Egan, I. M. Morrison and J. N. C. Whyte. Unpublished results.
- <sup>34</sup> G. O. Aspinall, K. Hunt, and I. M. Morrison. Unpublished results.
- <sup>35</sup> V. M. Parikh and J. K. N. Jones. *Canad. J. Chem.* **44**, 327, 1531 (1966).
- <sup>36</sup> E. L. Hirst, J. K. N. Jones, and W. O. Walder. *J. Chem. Soc.* 1225 (1947).

# THE CHEMICAL STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

ROGER W. JEANLOZ

*Laboratory for Carbohydrate Research, Harvard University Medical School,  
Massachusetts General Hospital, Boston, Massachusetts 02114, U.S.A.*

## I. INTRODUCTION

In the past, the role of organic chemistry has been to establish the structure of natural products, to duplicate these products, and synthesize infinite variations of them for the amelioration of our standard of living. Recently, however, organic chemistry, together with its derived disciplines biochemistry and molecular biology, has broadened in scope and has started to give us basic knowledge about biological processes. This knowledge is valid, however, only when it is based on chemical structures, the identification of which has been made by rigorous methods and eventually supported by synthetic proofs. This knowledge is valid, in other words, only when it is based on the same foundation as the chemistry of natural products has been based in the past.

Most developments in the field of molecular biology have concerned proteins and nucleic acids. There are, however, biological phenomena (for example, the immunological protection of the cell, the duplication of the bacterial cell, and its penetration by viruses) in which carbohydrate structures play a major role. When we come to know the chemical structure of the rigid and insoluble component that gives form and protection to bacteria, we will better understand some important biological phenomena at the molecular level. To take two particularly interesting examples, we will be able to explain in chemical terms the penetration of viruses into the bacterial cell and also the action of antibiotics, such as penicillin, which block the formation of the cell wall. Only recently, moreover, has the spatial structure of egg-white lysozyme, one of the enzymes that cause the lysis of bacteria, been elucidated<sup>1</sup>. When we know completely the chemical structure of the lysozyme substrate, we will be in a position to explain the mechanism of enzyme action.

In recent years, methods have been developed for the preparation of cell wall material devoid of cytoplasmic or membrane material, so that chemical methods may be applied to the purified material. Cell walls of gram-positive bacteria were found to have, in general, a simpler composition than do cell walls of gram-negative bacteria, which have complex protein-lipopolysaccharide antigens. We have to thank Salton, Work, Cummings, and Westphal, among others, for the pioneering work which determined the composition of bacterial cell walls.

One of the first cell-wall structures to be studied was that of *Micrococcus lysodeikticus*, which is the substrate traditionally used for the study of egg-white lysozyme. Taking advantage of the solubilization of the cell wall by the

ROGER W. JEANLOZ

enzyme, Salton and Ghysen<sup>2</sup> and, independently, Perkins<sup>3</sup> have isolated fragments of low molecular weight, for which chemical structures have been proposed. On the basis of this work, a general structure was suggested. Polysaccharide chains composed of alternating units of 2-acetamido-2-deoxy-D-glucose (*N*-acetylglucosamine) and 2-acetamido-3-*O*-(D-1-carboxyethyl)-2-deoxy-D-glucose (*N*-acetylmuramic acid) are linked to a peptide network composed of D- and L-alanine, D-glutamic acid and L-lysine residues. Subsequently, when a polysaccharide composed of D-glucose and 2-amino-2-deoxymannuronic acid was isolated from the same cell wall<sup>4</sup>, it was clear that even one of the relatively simple cell walls has a complex chemistry.

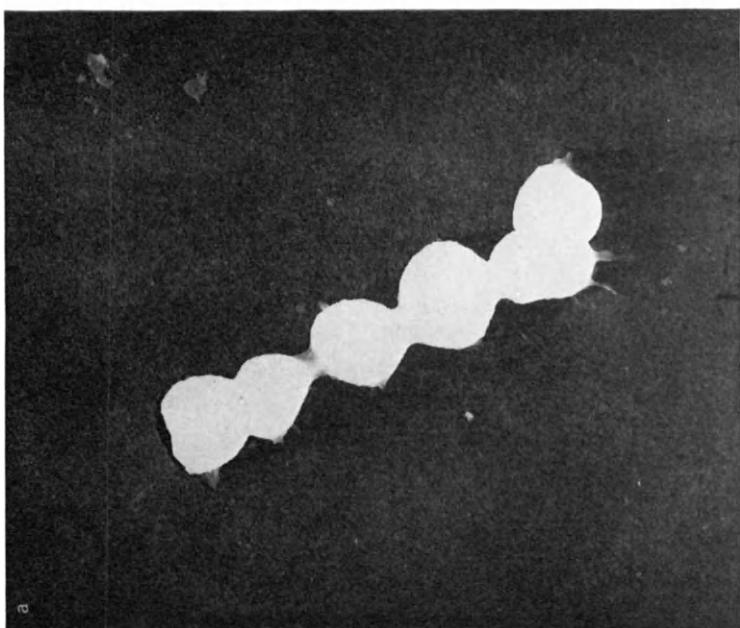
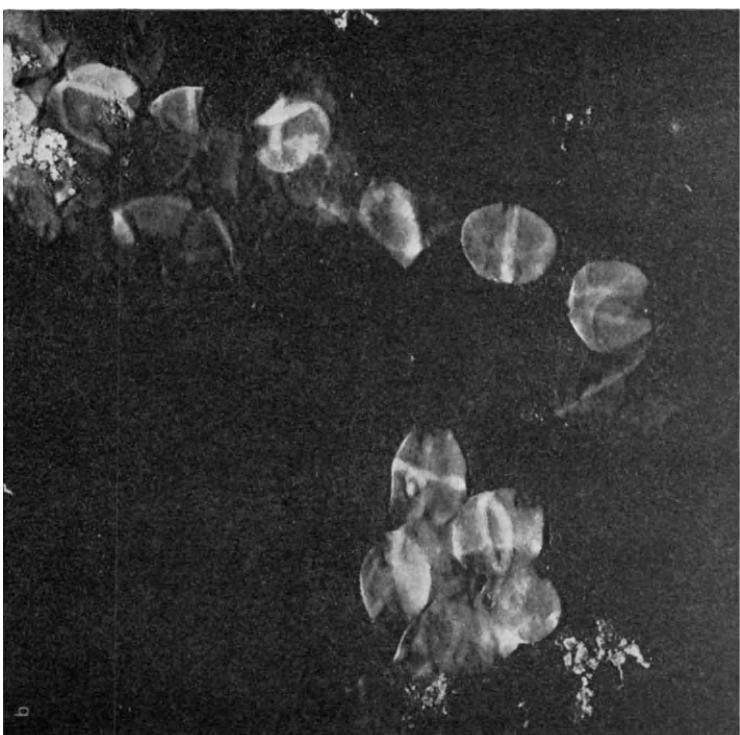
In the following pages, I will present the results of studies made in collaboration with Drs. Sharon, Flowers, Osawa, Nasir-ud-Din, Hoshino, Gross, Zehavi, Miss Walker, and Mrs. Jeanloz. These studies were carried out in order to solve some of the problems of carbohydrate chemistry presented by the structure of the cell wall of *M. lysodeikticus*. Specifically, I will discuss our attempt to determine the structure of the *N*-acetylglucosamine-*N*-acetylmuramic acid polysaccharide and its relation with the D-glucose and 2-amino-2-deoxymannuronic acid components; and then I will discuss the chemistry of muramic acid and the chemistry of synthetic substrates of lysozyme.

## II. ISOLATION, DEGRADATION, AND FRACTIONATION OF *M. LYSODEIKTICUS* CELL WALLS

The preliminary study<sup>2,3</sup> of the chemical structure of the cell wall of *M. lysodeikticus* had been based on degradation by lysozyme, followed by dialysis and isolation of the components by paper chromatography; structures were proposed for these components on the basis of colour reactions and periodate oxidation. In the present study the dialyzable and nondialyzable materials were obtained under conditions similar to those previously described. They were fractionated by adsorption on columns of acidic and basic resins, on DEAE-cellulose, or by precipitation with cetylpyridinium chloride (Figure 1).

In order to obtain sufficient amounts of material to be studied, a large-scale method of preparation was devised<sup>5</sup>. This method was based on the principle that the bacterial cell of *M. lysodeikticus* is disrupted by homogenization in the presence of glass beads. The resulting homogenate was purified by differential centrifugation, and the cell walls were further treated with trypsin. The material thus obtained is insoluble; its purity was controlled by electron microscopy (Figure 2) and, after degradation with lysozyme, by ultraviolet adsorption to show the absence of nucleic acids, which would indicate cytoplasmic material. Since the proportion of amino acid and carbohydrate components varies with the conditions of bacterial growth (and may also depend on the method by which the cell wall is prepared) various results have been published. A typical preparation shows the following relative proportions for each molecule of glutamic acid: 1 molecule of lysine, 1.5 molecules of glycine, 2.5 molecules of alanine, 1 molecule of 2-acetamido-2-deoxyglucose, slightly less than 1 molecule of 2-acetamido-3-*O*-(D-1-carboxyethyl)-2-deoxyglucose, 3 molecules of glucose, and an

Figure 2. Electron micrographs (shadowed with chromium) of: (left) *M. lysodeikticus* cells; and (right) of isolated *M. lysodeikticus* cell walls (Courtesy of Dr. J. Gross)



## STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

unknown proportion (but not more than 1 molecule) of 2-amino-2-deoxy-mannuronic acid<sup>6,8</sup>.

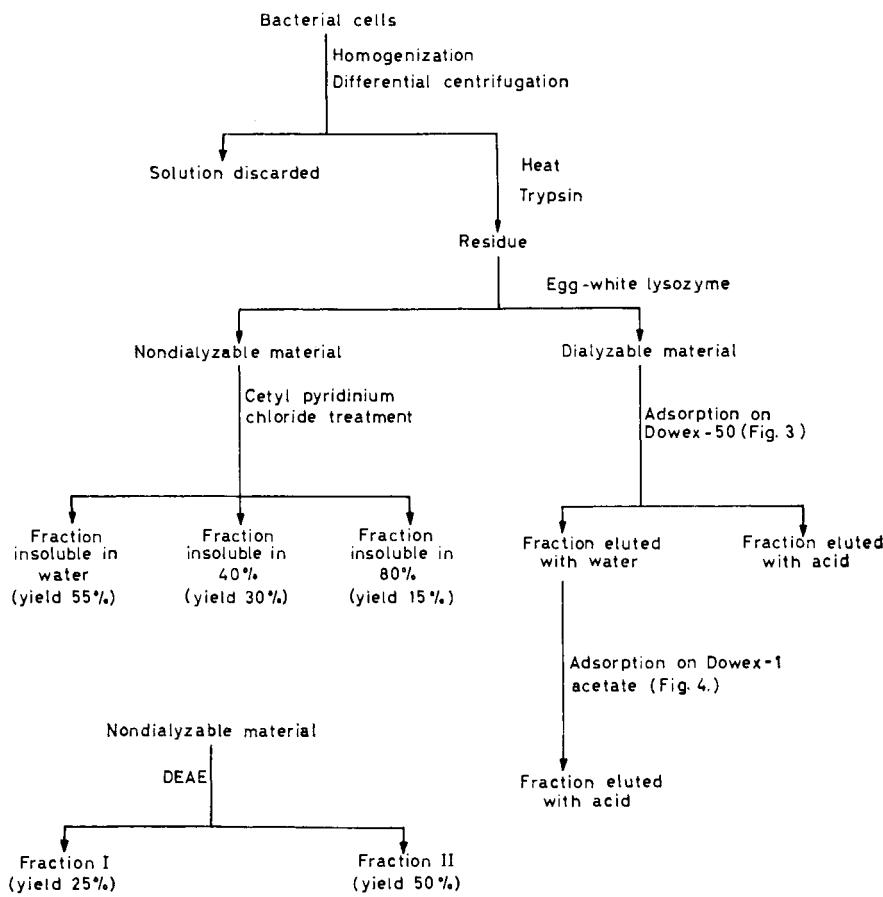


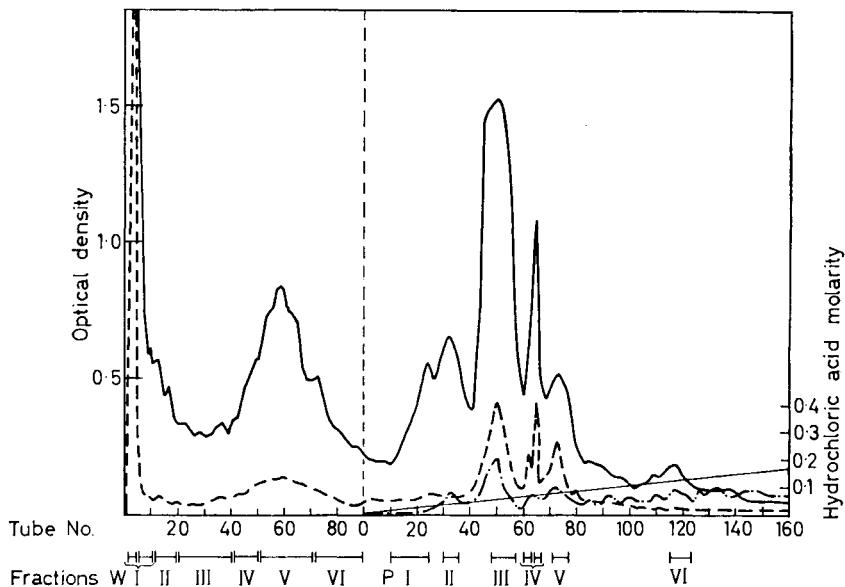
Figure 1. Scheme for the preparation of *M. lysodeikticus* cell walls, enzymatic degradation and fractionation

In previous studies of the chemical structure of the *M. lysodeikticus* cell wall degradation by lysozyme<sup>2,3</sup>, by acid hydrolysis<sup>9</sup>, and by methanolysis had been used<sup>10</sup>. Because of its selectivity, the first procedure is the most promising, although acid hydrolysis did confirm some of the results previously obtained. Methanolysis presented evidence for a covalent link between the glucose component and the glucosamine-muramic acid polysaccharide, as well as evidence for the linkage of glycine to D-glutamic acid.

Fractionation of large amounts of the dialyzable material which results from the action of lysozyme on the cell wall had been first performed by adsorption on charcoal and on Dowex-1 acetate columns; both processes were used successfully to isolate a disaccharide and a tetrasaccharide, composed of equimolecular amounts of glucosamine and muramic acid, which had been isolated earlier on a microscale. Recently, this technique was further refined<sup>11</sup>; as a preliminary step, the dialyzable portion of the

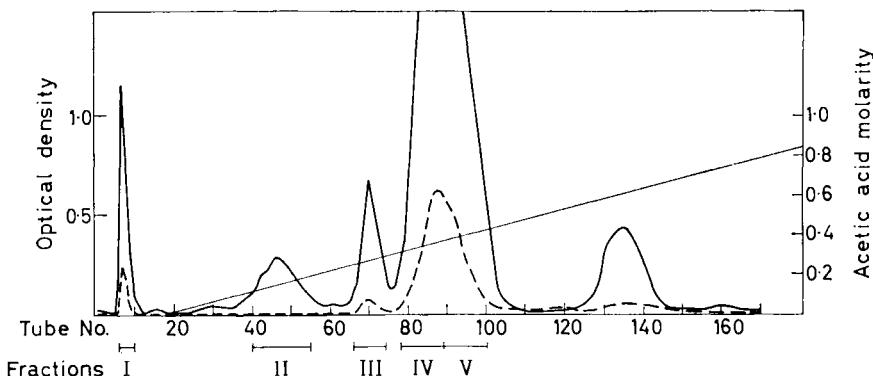
ROGER W. JEANLOZ

lysozyme hydrolyzate was adsorbed on Dowex-50. Elution was accomplished first with water and then by a gradient of hydrochloric acid (*Figure 3*). The peaks were determined by the Park-Johnson test<sup>12</sup> (reducing sugars),



*Figure 3.* Fractionation on Dowex-50 of the dialyzable fraction of *M. lysodeikticus* cell walls after lysozyme degradation. Fractions W-I to VI were obtained by elution with water, fractions P-I to VI by elution with a gradient of hydrochloric acid. (—) Park-Johnson test; (---) Morgan-Elson test; (- - -) ninhydrin test

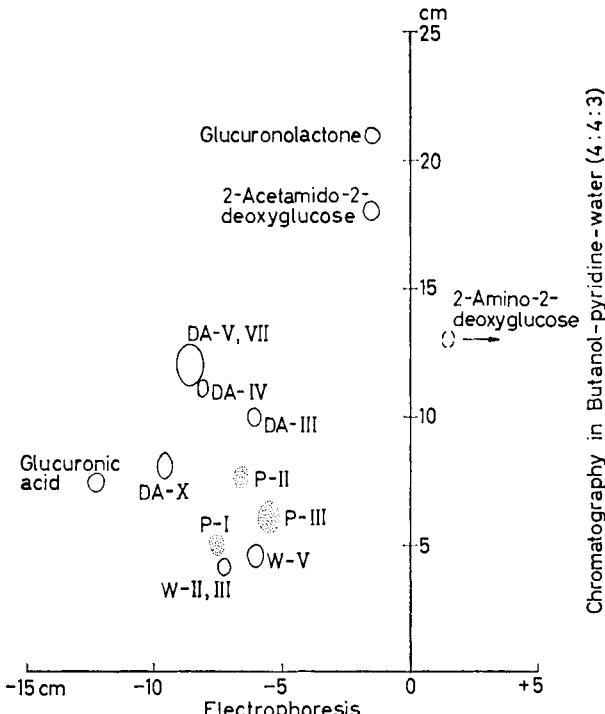
by the Morgan-Elson test<sup>13</sup> (acetamidodeoxy sugars), and by the ninhydrin test (amino acids). Each peak was investigated by paper chromatography in two solvent systems and by paper electrophoresis. The first peak, eluted with water, was further purified by adsorption on Dowex-1 ( $\text{CH}_3\text{COO}^-$ ), and then was eluted with a gradient of acetic acid (*Figure 4*); the resulting peaks were determined in the manner described above. The results of the



*Figure 4.* Fractionation on Dowex-1 acetate of the fraction W-I eluted from the Dowex-50 column. Fraction I was obtained by elution with water, and Fractions II to V by a gradient of acetic acid. (—) Park-Johnson test; (---) Morgan-Elson test

## STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

paper chromatography and of the electrophoresis were reported on a two-dimensional map (*Figure 5*): at least six substances reacted with the aniline phosphate reagent and at least three reacted with both the aniline phosphate and ninhydrin reagents.



*Figure 5.* Two-dimensional map of the spots obtained after paper electrophoresis and paper chromatography. All spots reacted with the aniline phosphate reagents. Spots corresponding to P-I, P-II, and P-III reacted, in addition, with the ninhydrin reagent

Since Perkins<sup>4</sup> had described the isolation, from *M. lysodeikticus* cell wall, of a polysaccharide composed of D-glucose and 2-amino-2-deoxymannuronic acid, attempts to fractionate the nondialyzable part of the *M. lysodeikticus* digest were based on adsorption on diethylaminoethylcellulose and on fractional precipitation with cetylpyridinium chloride<sup>8</sup>. Both methods have been used extensively for the purification of polysaccharides which contain uronic acid. In addition to the usual amino acids, the nondialyzable fraction contained 32 per cent glucose, 10 per cent glucosamine, 7 per cent muramic acid and an unknown (but no more than 15 per cent) percentage of 2-amino-2-deoxymannuronic acid; this last component, which is degraded in large part by acid hydrolysis, has never been determined quantitatively, since no method of measurement in presence of amino acids and muramic acid has been devised as yet. Fractionation with cetylpyridinium chloride gave three main fractions, one insoluble in water, the other two insoluble in ethanol solutions of 40–50 per cent and 80–90 per cent, respectively (*Figure 1*). Adsorption on diethylaminoethylcellulose, followed by elution with a phosphate buffer gradient at pH 6, gave two main fractions in yields of 25 per cent and 50 per cent.

Table 1. Properties of the fractions obtained from the nondialyzable fraction of *M. lysodeikticus* cell walls after lysozyme degradation

Fractions	[ $\alpha$ ] <sub>D</sub> in water (degrees)	N (%)	P (%)	Acetyl (%)	Hexoses (%)	Hexosamine (%)	Muramic Acid (%)	Amino Acids (%)			
								Glu	Ala	Gly	Lys
CPC-Insoluble in water	+31	7.2	1.1	20.1	32	10	7	6	7	10	8
CPC-Insoluble in 40% ethanol	+18	11.5	0.5	17.0	8	10	4	3	10	4	8
CPC-Insoluble in 80% ethanol	+40	4.9	2.7	17.9	45	8	5	3	4	2	2
DEAE-I	+22				23	10	7	+	+	+	
DEAE-II	+34				34	7	7	+	+	+	

## STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

No clear-cut difference was found among the various fractions obtained by cetylpyridinium chloride precipitation or diethylaminoethylcellulose adsorption. Glass-fibre electrophoresis in phosphate buffer at pH 6.2 showed the presence of two components, but sedimentation analysis showed only one homogeneous peak. Optical rotation and quantitative determination of the components showed the two main fractions, obtained from each procedure of separation, to be quite similar (*Table I*); consequently the fraction giving a cetylpyridinium complex insoluble in water was investigated further.

### III. CHEMICAL STRUCTURE OF ISOLATED FRAGMENTS

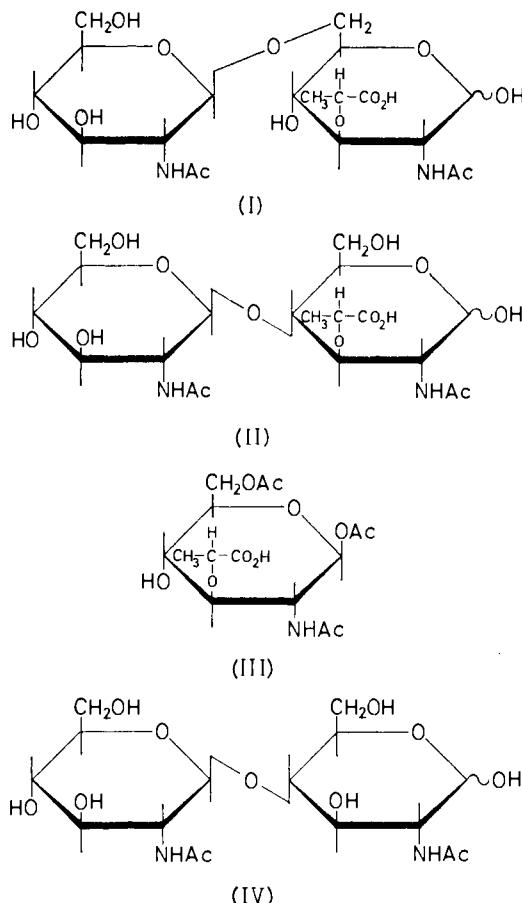
#### A. Dialyzable component

The water eluate of the Dowex-50 column could be separated into four main fractions on Dowex-1 ( $\text{CH}_3\text{COO}^-$ ). Two of these fractions correspond to the di- and tetrasaccharide previously isolated (DA-V-VII and DA-X respectively). The two other fractions correspond to a disaccharide (DA-IV) and a tetrasaccharide (DA-III), as was determined from the reducing properties and from the speed of migration on paper electrophoresis. The four components gave, after hydrolysis, 2-amino-2-deoxyglucose. In addition, upon paper electrophoresis, the first two components showed a spot corresponding to 2-amino-3-*O*-(D-1-carboxyethyl)-2-deoxyglucose (muramic acid). However, when the last two components were tested, the spot did not move quite as fast as the one produced by muramic acid. Furthermore, the spot produced by the acidic component of the two last-mentioned substances gave a much weaker reaction with the alkaline silver nitrate reagent, a characteristic of sugars having the *manno*- or *talo*- configuration. The colour reaction shown by this acidic component, after treatment with ninhydrin, did not correspond to that given by 2-amino-2-deoxymannuronic acid. Further investigations to elucidate the structure of this component are in progress.

##### 1. Structure of disaccharide DA-V-VII

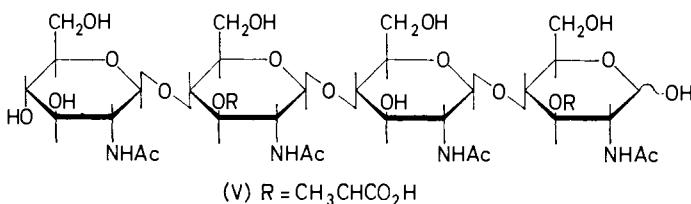
Structure (I) was proposed for the disaccharide DA-V-VII by earlier investigators<sup>2,3</sup>, who relied mainly on periodate oxidation and colour reactions. Later, the synthesis of this disaccharide was achieved<sup>12-14</sup>. Since a comparison of the crystalline, fully-acetylated methyl ester of the natural product with that of the synthetic product has shown that the two are not identical, the structure of *O*-2-acetamido-2-deoxy- $\beta$ -D-glucopyranosyl-(1 → 4)-2-acetamido-3-*O*-(D-1-carboxyethyl)-2-deoxy-D-glucose (II) has been proposed for the natural disaccharide<sup>7,15</sup>.

The synthesis of this disaccharide by condensation of 2-acetamido-1, 6-di-*O*-acetyl-3-*O*-[D-1-(methyl carboxylate)ethyl]- $\beta$ -D-glucopyranose (III) with 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy- $\alpha$ -D-glucopyranosyl bromide, or by starting from di-*N*-acetylchitobiose (IV) and adding the lactyl side chain, was attempted, but has not been successful as yet.



## 2. Structure of tetrasaccharide DA-X

Structure (V) has been proposed for the tetrasaccharide DA-X. The proposal was based on the observation that when lysozyme splits the tetrasaccharide, not only do some higher molecular oligosaccharides result from transglycosylation, but also two molecules of the disaccharide described



in the preceding paragraph are produced<sup>18</sup>. The linkage between C-1 of the non-reducing 2-acetamido-3- $\bar{\beta}$ -(D-1-carboxyethyl)-2-deoxy- $\beta$ -D-glucopyranosyl residue and the 2-acetamido-2-deoxy- $\beta$ -D-glucopyranosyl residue has been assumed to be (1 → 4), since lysozyme attacks oligosaccharide derived

## STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

from chitin. Some additional evidence for this linkage was gained from the results of the periodate oxidation of a trisaccharide obtained by enzymatic degradation of the tetrasaccharide<sup>17</sup>. Moreover, the hydrolyzate of the methylated nondialyzable fraction gave, in large proportion, the 3,6-dimethyl-ether of 2-amino-2-deoxy-D-glucose (*see* next paragraph), which is a further indication for a (1 → 4) linkage.

### B. Nondialyzable component

The fraction precipitated by cetylpyridinium chloride in water solution presents a composition very similar to that of the main fraction eluted from the diethylaminoethylcellulose column; it was, therefore, investigated<sup>8</sup> further.

Attempts to separate a polysaccharide composed of glucose and 2-amino-2-deoxymannuronic acid by treatment according to the method of Perkins<sup>4</sup>, with trichloroacetic acid for 48 hours at 35°, was not successful. The result was a peptidoglycan, isolated in a yield of 97 per cent, which still contained 31 per cent glucose, 11 per cent glucosamine, and 9 per cent muramic acid; this substance was not degraded further by treatment with lysozyme.

In order to ascertain the presence of 2-amino-2-deoxy-mannuronic acid, the fraction investigated was treated with diborane in diglyme solution; the reduced product showed, after hydrolysis, the presence of 2-amino-2-deoxymannose.

#### 1. Periodate oxidation

The peptidoglycan obtained from the water-insoluble cetylpyridinium complex was oxidized with excess periodate and then reduced with sodium borohydride. It was further treated with dilute acid and dialyzed. The dialyzate showed the presence of glycerol, and the remaining nondialyzable peptidoglycan contained 2 per cent glucose, 9 per cent glucosamine, and 14 per cent muramic acid. A second periodate oxidation removed the remaining glucose units.

#### 2. Methylation

The peptidoglycan was methylated with dimethyl sulphate and sodium hydroxide in an atmosphere of nitrogen at low temperature. The process was repeated until the content in methoxyl groups reached a limit at 18.1 per cent; the infrared spectrum indicated complete methylation. After acid hydrolysis and removal of the acid, the hydrolyzate was fractionated on a Dowex-50 column with a gradient of hydrochloric acid.

The methylated sugars were identified by a variety of methods: paper chromatography in four different solvent systems, paper electrophoresis, periodate oxidation of the sugar (or of its glycitol derivative) followed by paper chromatography, gas-liquid chromatography of the methyl glycoside and of its trimethylsilyl derivative, degradation of the 2-amino-2-deoxy sugars with ninhydrin and, finally, crystallization of the sugars (or of the

ROGER W. JEANLOZ

azoyl derivative of the corresponding glycitol). The results are reported in *Table 2*.

*Table 2.* *O*-Methyl derivatives isolated from the hydrolysate of a methylated nondialyzable fraction (CPC-Complex insoluble in water)

Starting material:	1.37 g
Methylated material recovered:	0.98 g ( $\text{OCH}_3$ : 18%)
Methylated material hydrolysed:	0.46 g
<i>O</i> -Methyl derivatives isolated:	3- <i>O</i> -Methyl- $\text{D}$ -glucose (16 mg) 2,3-Di- <i>O</i> -methyl- $\text{D}$ -glucose (32 mg) 2,3,4-Tri- <i>O</i> -methyl- $\text{D}$ -glucose (50 mg) 2-Amino-2-deoxy-3- <i>O</i> -methyl- $\text{D}$ -glucose (14 mg) 2-Amino-2-deoxy-3,6-di- <i>O</i> -methyl- $\text{D}$ -glucose (29 mg)

The methylated sugars isolated in order of decreasing amounts are as follows: the 2,3,6-trimethyl, 2,3-dimethyl and 3-methyl ethers of  $\text{D}$ -glucose for the neutral sugars; and the 3,6-dimethyl and 3-methyl derivatives of 2-amino-2-deoxy- $\text{D}$ -glucose for the amino sugars.

#### IV. TENTATIVE CHEMICAL STRUCTURE OF THE CARBOHYDRATE MOIETY

The carbohydrate components of the *M. lysodeikticus* cell wall identified up to the present time are 2-amino-2-deoxy- $\text{D}$ -glucose ( $\text{D}$ -glucosamine), 2-amino-3-*O*-( $\text{D}$ -1-carboxyethyl)-2-deoxy- $\text{D}$ -glucose (muramic acid),  $\text{D}$ -glucose and 2-amino-2-deoxymannuronic acid. The possibility that other unidentified sugars exist in this particular cell wall cannot be excluded; one of them may be a compound very similar to muramic acid.

As a result of the work of Salton, Perkins, and their associates, it has been assumed that the first two compounds and the last two form two separate polysaccharides. In view of the work described in the preceding paragraphs, this hypothesis should be modified. While it is not as yet possible to ascertain the position of the 2-amino-2-deoxymannuronic acid component, it seems most likely that the  $\text{D}$ -glucose units are covalently linked with the peptidoglycan composed of glucosamine and muramic acid.

There is no reason to doubt the existence of a glycan chain consisting of alternating units of *N*-acetylglucosamine and *N*-acetylmuramic acid, an hypothesis suggested by earlier investigators. However, until such a chain has been isolated, devoid of peptide side-chains, until its molecular weight has been determined, and until its complete chemical structure has been elucidated, it will not be possible to ascertain the role of this glycan in the structure of the cell wall.

The 2-acetamido-2-deoxy- $\beta$ - $\text{D}$ -glucopyranosyl and 2-acetamido-3-*O*-( $\text{D}$ -1-carboxyethyl)-2-deoxy- $\beta$ - $\text{D}$ -glucopyranosyl units are linked at C-4, as is suggested by studies of the dialyzable di- and tetrasaccharides, and by the fact that the 3,6-dimethyl ether of 2-amino-2-deoxy- $\text{D}$ -glucose has been isolated in the largest proportion from the methylated derivative of one of the nondialyzable fragments (*Table 2*). The amount of 3-methyl ether of 2-amino-2-deoxy- $\text{D}$ -glucose that has been isolated from the hydrolysate of the methylated nondialyzable fraction is too large to be discounted as a result

## STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

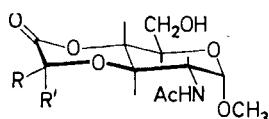
of incomplete methylation; its presence may indicate branching along the chain of the peptidoglycan.

Perkins has shown, by the isolation (in a very small yield) of a polysaccharide consisting of glucose and 2-amino-2-deoxymannuronic acid, that these two components are linked together. The results of the methylation, in agreement with those of the periodate oxidation, have established that the D-glucose units are linked at C-6 and that none of them is located at the terminal, non-reducing end of a chain. It is quite logical to assume that they are the first components of a branch chain attached to the chain of the peptidoglycan and that they are linked only to the glucosamine residues; but there is no direct evidence as yet to support this suggestion. Furthermore, it is not known whether the chains starting with the D-glucose units consist of alternating units of D-glucose and 2-amino-2-deoxymannuronic acid components; nor is it known in which position the 2-amino-2-deoxymannuronic acid units are linked. The large proportion of 3-methyl and 2,3-dimethyl ethers of D-glucose suggest extensive ramification of the side-chains. Much work remains to be done until a chemical structure of the whole bacterial cell wall can be presented.

## V. PRESENT STATUS OF MURAMIC ACID CHEMISTRY

The chemistry of 2-amino-3-O-(D-1-carboxyethyl)-2-deoxy-D-glucose (muramic acid), a component found in all bacterial cell walls has been the subject of a few recent reviews<sup>18-21</sup>. Some of the problems concerning the structure of the bacterial cell wall which have been investigated recently in our laboratory include the following: the configuration of the side-chain; new methods of synthesizing muramic acid; and the preparation of derivatives which may serve as substrates for lysozyme (such as the  $\beta$ -D-glycosides), or which may be used to elucidate the structure of the peptidoglycan chain (such as the methyl ethers).

The D-isomery at C-2 of the side chain of muramic acid was assigned after comparison of the optical rotation of muramic acid with that of optically active ethers of D-lactic acid<sup>22</sup> and after consideration of the mechanism of the reaction assumed to take place in the synthesis starting with L-chloropropionic acid<sup>23</sup>. These hypotheses are, however, open to criticism; we are attempting to prove (or disprove) the proposed configuration by using different methods. The preparation of the lactone (VI) {m.p. 246-8°,  $[\alpha]_D + 227^\circ$  in acetone}



(VI) R = H, R' = CH<sub>3</sub>

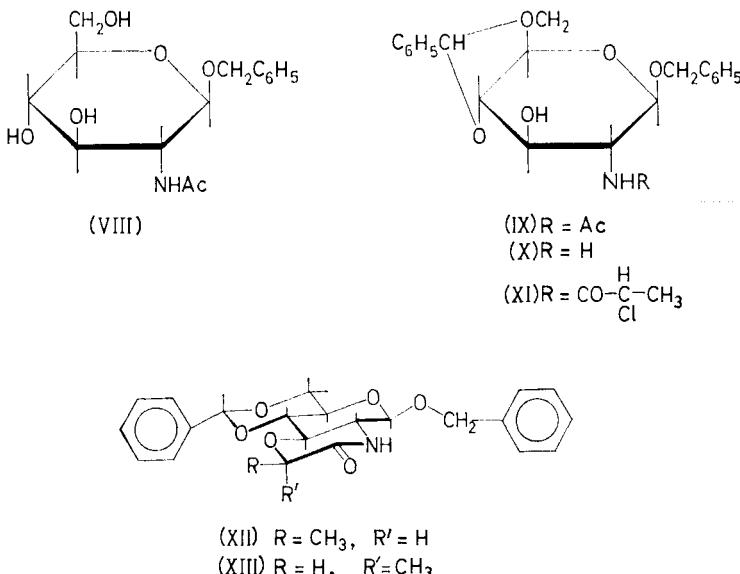
(VII) R = CH<sub>3</sub>, R' = H

has been reported previously<sup>24</sup>. A similar lactone (VII), which differs solely in the isomery of the lactyl chain, has now been prepared {m.p. 254-6°,  $[\alpha]_D + 168^\circ$  in acetone}<sup>25</sup>. When sufficient amounts of both lactones are available, determination of their optical rotatory dispersion and

ROGER W. JEANLOZ

of the nuclear magnetic resonance spectra may give additional information on the conformation of the methyl group of the lactyl side chain.

Preparation of a disaccharide composed of glucosamine and muramic acid, in which the latter component is linked at C-4, has not as yet been accomplished. The bulky lactic acid side-chain seems to present a major obstacle, since an attempt to prepare the 4-benzoate of the 6-trityl ether was not successful. In order to study less hindered intermediates, a lactam derivative has been prepared through the sequences (VIII) to (XII)<sup>26</sup>.



Study of the optical rotatory dispersion and nuclear magnetic resonance spectra of the lactam (XII) and of the isomeric lactam (XIII) formed during the synthesis may also establish the configuration of the lactyl side chain.

## VI. LYSOZYME SUBSTRATES

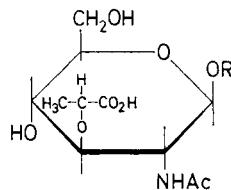
When the site of action of egg-white lysozyme in the bacterial cell wall was elucidated and when the spatial structure of the enzyme was determined, there was added incentive to prepare synthetic substrates.

In order to study the specificity of egg-white lysozyme and its interaction with muramic acid derivatives, both the *p*-nitrophenyl (XIV) and benzyl  $\beta$ -D-glycosides (XV) have been synthesized<sup>27</sup>. Both products have been shown to be resistant to enzymatic degradation; similar results have been reported for other  $\beta$ -D-glycosides of muramic acid<sup>19</sup>.

Many years ago, degradation products of chitin were known to be substrate for egg-white lysozyme; consequently interest has centred on di-N-acetyl-chitobiose. Thus, recently, the *p*-nitrophenyl  $\beta$ -D-glycoside of the disaccharide (XVI) has been synthesized and shown to be a reactive substrate for the enzyme<sup>28</sup>. However, the large proportion of enzyme necessary to obtain a rate sufficient to be measured raises some doubt on the validity of the

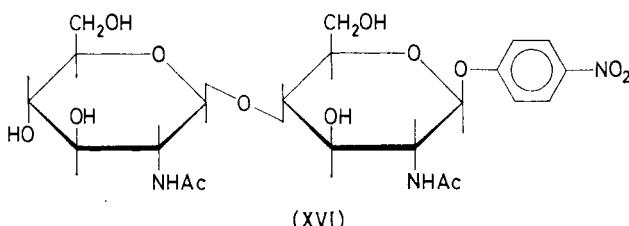
## STRUCTURE OF THE CELL WALL OF GRAM-POSITIVE BACTERIA

assumption that in a bacterial cell wall the linkage split by the enzyme is of the same type as in chitin [ $\beta$ -D (1  $\rightarrow$  4)]. Work now in progress on the synthesis of various oligosaccharides of 2-acetamido-2-deoxy-D-glucose will



(XIV) R =  $C_6H_5NO_2-p$

(XV) R =  $CH_2C_6H_5$



(XVI)

help to solve this problem, and will give further information on the structure of the bacterial cell wall.

## CONCLUSION

The problems I have briefly discussed show how complex a task it is to describe in chemical terms the organization of just one biological component (the cell wall) of rather simple organisms (bacteria). The results are, nevertheless, encouraging, because they show that the tools for such studies are available. But it is clear, at the same time, that many of the present procedures must be performed automatically if the desired solutions are to be found in a not too distant future.

## References

- <sup>1</sup> C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. *Nature* **206**, 757 (1965).
- <sup>2</sup> M. R. J. Salton. *Biochim. Biophys. Acta* **22**, 495 (1956); **34**, 308 (1959); J. M. Ghysen and M. R. J. Salton. *Biochim. Biophys. Acta* **40**, 462 (1960).
- <sup>3</sup> H. R. Perkins. *Biochem. J.* **74**, 186 (1960).
- <sup>4</sup> H. R. Perkins. *Biochem. J.* **86**, 475 (1963).
- <sup>5</sup> N. Sharon and R. W. Jeanloz. *Experientia* **20**, 253 (1964).
- <sup>6</sup> M. R. J. Salton. *The Bacterial Cell Wall*. Elsevier Publishing Corp., Amsterdam, 1964, p.274.
- <sup>7</sup> N. Sharon, T. Osawa, H. M. Flowers, and R. W. Jeanloz. *J. Biol. Chem.* **241**, 223 (1966).
- <sup>8</sup> Nasir-ud-Din and R. W. Jeanloz. Unpublished results.
- <sup>9</sup> H. R. Perkins and H. J. Rogers. *Biochem. J.* **69**, 15P (1958); **72**, 647 (1959).
- <sup>10</sup> R. W. Jeanloz and L. Dobbs. Unpublished results.
- <sup>11</sup> O. Hoshino and R. W. Jeanloz. Unpublished results.
- <sup>12</sup> H. M. Flowers and R. W. Jeanloz. *J. Org. Chem.* **28**, 1564 (1963).

ROGER W. JEANLOZ

- <sup>13</sup> H. M. Flowers and R. W. Jeanloz. *J. Org. Chem.* **28**, 2983 (1963).
- <sup>14</sup> T. Osawa and R. W. Jeanloz. *Carbohydrate Res.* **1**, 181 (1965).
- <sup>15</sup> R. W. Jeanloz, N. Sharon, and H. M. Flowers. *Biochem. Biophys. Res. Commun.* **13**, 20 (1963).
- <sup>16</sup> N. Sharon and S. Seifter. *J. Biol. Chem.* **239**, PC 2398 (1964).
- <sup>17</sup> M. Leyh-Bouille and D. J. Tipper. *Bull. Soc. Chim. Biol.* **47**, 1905 (1965).
- <sup>18</sup> G. Blix and R. W. Jeanloz. In *Amino Sugars Vol IA* (R. W. Jeanloz Ed.), Academic Press, New York. In the press.
- <sup>19</sup> R. W. Jeanloz. *Exposés Ann. Biochim. Med.* **27**, 45 (1966).
- <sup>20</sup> A. Gottschalk. *The Chemistry and Biology of Sialic Acids and related Substances*. Cambridge University Press, 1960.
- <sup>21</sup> M. R. J. Salton. *The Bacterial Cell Wall*. Elsevier Publishing Corp., Amsterdam, 1964, p.114.
- <sup>22</sup> R. E. Strange and L. H. Kent. *Biochem. J.* **71**, 333 (1959).
- <sup>23</sup> Y. Matsushima and J. T. Park. *J. Org. Chem.* **27**, 3581 (1962).
- <sup>24</sup> R. W. Jeanloz, H. M. Flowers, and E. Walker. *Abst. Papers Am. Chem. Soc.* 145th Meeting, New York, 1963, p. 18D.
- <sup>25</sup> R. W. Jeanloz, E. Walker, and D. A. Jeanloz. Unpublished results.
- <sup>26</sup> P. H. Gross and R. W. Jeanloz. *Abst. Papers Am. Chem. Soc.* Winter Meeting, Phoenix, 1966, C-32.
- <sup>27</sup> R. W. Jeanloz and E. Walker. Unpublished results.
- <sup>28</sup> T. Osawa. *Carbohydrate Res.* **1**, 435 (1966).

## **SECTION 2: STRUCTURAL ELUCIDATION OF NATURAL PRODUCTS OTHER THAN MACROMOLECULAR COMPOUNDS**

### **INTRODUCTORY REMARKS BY THE HONORARY PRESIDENT OF THE SECTION**

ROGER ADAMS

It was eight to ten years ago when initial steps were taken to organize a symposium on the chemistry of natural products. In the interim three successful symposia have been held with the fourth one now under way. During these years an ever increasing number of communications, describing the elucidation of the structure and syntheses of many complex products occurring in plant and animal life, have appeared. I intend to reminisce a bit this afternoon and to note the basic contributions of the last sixty or seventy years that have made possible the current brilliant successes.

But first I must cite conditions that existed in earlier times and a few examples of the achievements. As a student I had the privilege to study in 1912-1913 in the laboratory of Emil Fischer and later in that of Richard Willstaetter. Consequently I became acquainted with natural product researches of those days and of the preceding two decades. It is difficult for chemists trained after World War II to imagine how the structure of complex molecules could be solved almost exclusively by chemical methods. Yet at the turn of the century they were solved in that way, if you are willing to overlook absolute configuration. It often required years of painstaking research by large groups of chemists to complete the study of a single natural product.

An investigation was always started by detection of the presence of the functional groups by means of various reagents. Degradation experiments followed and the fragments separated. Each fragment was then studied chemically until it was identified or degraded further into identifiable products. With this information, assuming rearrangements during degradation did not distort the picture, the original molecule could usually be constructed. One of the handicaps during that period was the requirement of relatively large amounts of material for the numerous reactions and degradations; and in addition 0.1 to 0.2 grams were needed for each elemental analysis. This made it imperative to select for study readily available products, such as quinine, morphine, cocaine, strychnine, and brucine.

ROGER ADAMS

Yet without the many advantages of today, the investigations by the leaders in organic chemistry were noteworthy. I cite, for example, the determination of structure of many of the tropin alkaloids and subsequent synthesis by Willstaetter and assistants. Their publications on the anthocyanins, the colouring matter of flowers and fruits, represent some of the finest aspects of chemistry found anywhere in the literature. I recall that just adjacent to the laboratory there were acres of red geraniums and of cornflowers from which each morning the dieners would pluck fresh petals and promptly bring them to the laboratory for immediate extraction. These investigations extended over many years. I have often wondered how long it would take the clever chemist of today with modern facilities to realize the same goals. And this applies to the natural product studies of the following three decades.

Molecules which contained many different types of rings presented much more difficulty to the investigator who had to rely merely on chemical methods. Satisfactory degradation reactions were more difficult to discover, for frequently they involved rearrangements or often complete destruction; this was especially true of molecules composed exclusively of heterocyclic rings. For example, I may refer to the many dozens of papers published by Leuchs and his coworkers on the structure of brucine and strychnine as well as a long list on the same alkaloids by Robinson and associates before structural formulae were proposed adequate to explain the chemical reactions. Even so, the structures have been modified more recently when modern techniques made more detailed investigation possible.

It must not be overlooked also that in those early days no satisfactory procedure was available for purification of natural products contaminated with materials closely related physically and chemically. Chromatography had not been developed. Months were often consumed in a preliminary investigation to isolate a substance of adequate purity for structural study. Physical tools were not available excepting polarimeters and refractometers. Equipment for determining absorption spectra in the visible and ultraviolet was primitive and rarely used by the organic chemist. It was primarily of interest to the physical chemist who measured and reported oscillation frequencies and absorption wavelengths usually of impure samples of material. It was years afterwards before ultraviolet absorption equipment was developed suitable for effective utilization by the organic chemist.

The first important contribution which aided the natural product chemist was the introduction of microanalyses during the 1920s. This made it possible to function without the amount of material previously considered essential. The development of laboratory apparatus suitable for running reactions with small quantities of reagents followed.

The second contribution came in the late 1920s and 1930s with the gradual perfection of chromatographic procedures for separation of mixtures of compounds closely related physically and chemically. To be sure, these methods were not what they are today but they were very serviceable. Satisfactory ultraviolet spectrophotometers also appeared during this period and proved of real significance to the chemist particularly when aromatic nuclei and unsaturated groups were present in the molecule.

But the most salient breakthrough for the natural product investigator

## INTRODUCTORY REMARKS

came after World War II with the perfection of infrared spectrophotometers which furnished precise information concerning functional groups. Then came nuclear magnetic resonance equipment which made possible further deductions concerning atoms and groups and, more importantly, their relative configurations. With n.m.r. to supplement i.r., particularly since i.r. sometimes fails to identify functional groups in complicated structures, the chemist was supplied with a spectacular means for accurate insight into the constitution of a molecule. Simultaneously, x-ray crystal structure techniques proved their value, and unique chromatographic and counter-current procedures for separation of molecules resistant to orthodox methods were added to the list. Then came equipment for optical rotatory dispersion, electron spin resonance, and finally for analytical and mass spectrometry to aid further in solving the chemist's problems.

The natural product investigator, however, must have also a comprehensive knowledge of synthesis. The chemist primarily interested in synthesis now has a reservoir of techniques and reagents not available in earlier times. Many new named reactions of general applicability as well as a host of others of more specific utility have been discovered in the last thirty to forty years. The number of organic chemicals and solvents commercially available has multiplied manifold. Dozens of reagents have been uncovered for modifying and protecting functional groups, many suitable for reactions in molecules containing other sensitive points of attack. The present knowledge of rearrangements prone to certain types of molecules often prevents the chemist from being misled. New catalysts with very specific action have been placed at the disposal of the investigator. Almost any complex natural product, excluding macromolecules, may now be synthesized. However, what is needed for success is a genius to conduct the research, a man who is omniscient in chemistry, who has absorbed and digested a library of information on reactions, reagents, and techniques and who can skillfully apply this knowledge.

The natural product chemist in earliest days required positive identification of small degradation fragments to solve his problems. In later years he was able to accelerate his progress by the use of certain physical methods. Today, the time has been reached when the elucidation of structure is usually initiated by physical methods and supplemented by chemical studies. Indeed, when very small quantities of material are available, confirmation of the structure established through physical methods is often attempted by synthesis. And today the chemist does not consider his problem completed without the knowledge of the absolute configuration of the atoms and groups.

I have cited previously a few of the researches of former chemists. Now I must mention one or two of the outstanding researches of the recent past to exemplify what may now be accomplished.

The colouring matter in black tea formed probably by an enzyme-catalyzed process during the fermentation of green to black tea is known as theaflavin. Initial chemical studies by Takino, followed by an examination of the u.v., i.r., n.m.r., analytical and mass spectra by Gianturco, Takino, and associates, led to a tentative structure consisting of two flavylin residues attached to a benzotropolone nucleus. They then proceeded by a neat synthesis, involving

### ROGER ADAMS

only a very few steps, to obtain the natural product. A few months later Ollis and his associates, following an extensive study of theaflavin, established that the tentative structure proposed by the other workers was incorrect only in the configuration of the bonds attaching the flavinyl nuclei to the benzotropolone.

For a most elegant synthesis reported recently I refer to the work on cephalosporin by Woodward and associates in the CIBA laboratories in Basle. By a short series of most ingeniously devised reactions the natural product was obtained.

The structural information gained from various spectra is so revealing that a new vista is opened to the natural product chemist. He is now able with extremely small quantities of material, usually obtained by one of the invaluable chromatographic procedures, to characterize almost any substance.

Results in the natural product field have already changed the pattern of thinking of the biochemist, physiologist, botanist, zoologist, entomologist, and microbiologist since there is now the possibility of isolating and identifying the critical chemical contributors to life processes.

Looking into the future, I can see only unlimited potentialities in the study of natural products. My view may be expressed by a taxi cab driver's homely but incisive reply to his customer. A visitor to Washington was riding in this man's cab. As they passed by the government building on which is found the inscription, "The past is prologue", the visitor turned to the driver and asked, "What does that mean?" The taxi driver replied, "Oh, that just means, you ain't seen nothing yet."

And now I have another illustration of recent brilliant experimentation. It involves the contributions of our speaker, Professor Karlson. While working as an assistant to Professor Butenandt in the Max Planck Physiological Institute in München he isolated the insect moulting hormone, ecdysone. He continued his researches at the same institute until recently, when he became professor at the Physiological Chemical Institute in Marburg. He has also found the time in recent years to contribute a superb text book on biochemistry which has been very widely accepted.

During the last twelve years, following the discovery of ecdysone, his researches have involved the biochemistry and physiology of the compound and a comprehensive chemical and physicochemical study that resulted in the establishment of its constitution. Only an investigator with a very astute mind, indefatigable patience and determination could undertake a programme of this kind and bring it to a successful conclusion. The reported structure has indeed now been confirmed by synthesis in two independent laboratories. Professor Karlson has pioneered the way for future investigations in this fascinating and intriguing field of insect hormones.

This afternoon his address will be on the "Chemistry of Insect Hormones and Insect Pheromones". It is a privilege to present Professor Karlson.

# THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

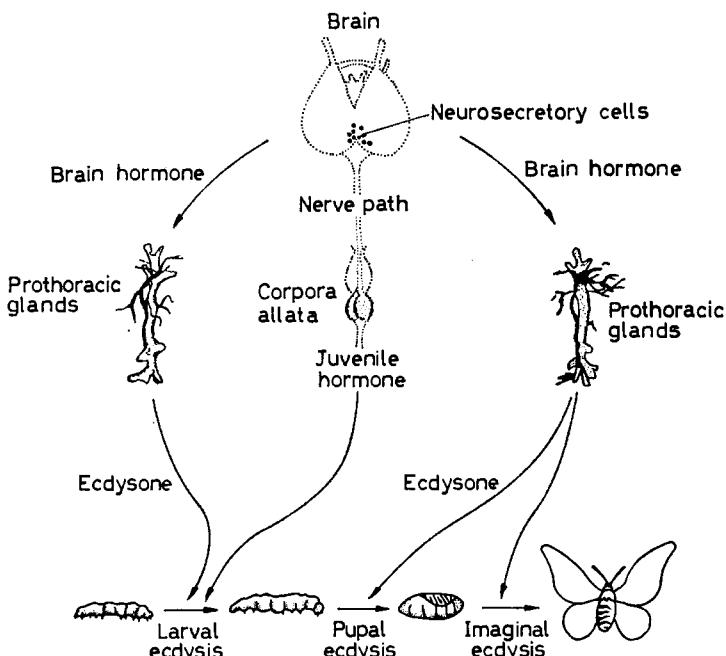
PETER KARLSON

*Physiologisch-Chemisches Institut der Philipps-Universität, Marburg, W. Germany*

## I. INSECT HORMONES

### 1. Introduction

While the hormones of vertebrates have been known to biologists for more than a century, and to chemists for well over half a century, the existence of insect hormones was recognized only thirty years ago, when Wigglesworth published his first work on *Rhodnius* and Bounhiol and Kühn repeated and confirmed the older, nearly forgotten work of Stefan Kopeč. (For a review of the earlier literature see references 1-6.)



*Figure 1.* Hormonal control of insect development. Three hormones are involved: the brain hormone, acting on the prothoracic glands, the juvenile hormone, secreted by the corpora allata, and ecdysone, secreted by the prothoracic glands. The bottom row shows the development from the caterpillar through the pupa to the moth. The larval ecdises are controlled by ecdysone and juvenile hormone, but the pupal and imaginal ecdises are induced only by ecdysone

Insect hormones are primarily involved in the regulation of post-embryonic development, which is characterized by moulting and metamorphosis. Larval moults are initiated by a secretion from the neurosecretory

PETER KARLSON

cells of the brain; this hormone stimulates the prothoracic glands to produce another hormone, now known as ecdysone, which acts on the periphery and causes moulting. Simultaneously, the corpora allata secrete the so-called "juvenile hormone" which guarantees the larval character of the moult. If juvenile hormone is lacking, the larvae will undergo a pupal or imaginal moult. *Figure 1* summarizes the hormonal control of development for the lepidoptera.

## 2. Chemistry of ecdysone

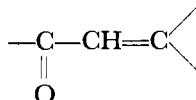
Our own work has been concerned with the prothoracic gland hormone, ecdysone<sup>7</sup>. Preliminary work on the purification of this hormone was done by Becker<sup>8</sup> in the laboratory of A. Kühn. We continued this work, using the *Calliphora* bioassay of Fraenkel<sup>9</sup> for tracing the biological activity, and used pupae of the commercial silkworm, *Bombyx mori*, for our extractions. It soon became clear that a large amount of starting material would be needed to obtain the pure hormone, and in 1953 the institute bought about 1200 kg of silk cocoons for the work on the prothoracic gland hormone as well as for the sex attractant to be discussed later. The male pupae served as starting material for our extraction, the first steps of which were done in the factory of Hoffmann-La Roche and Co. The concentrate was further purified in our laboratory, and in the spring of 1954 we isolated 25 mg of the crystalline hormone from the 500 kg pupae extracted<sup>10</sup>. This means a 20 million-fold purification. We can now estimate that the yield was not so bad; we got about 50 per cent of the hormone originally present in the extract. (The estimate is based on determinations of ecdysone content in *Bombyx* by Shaaya and Karlson<sup>11</sup>.)

Soon after the isolation of ecdysone, we detected in the extracts a second substance with biological activity. This substance was isolated and provisionally termed " $\beta$ -ecdysone"<sup>12</sup>; the amount available was so small (2.5 mg), that it could only be characterized spectroscopically<sup>2</sup>. Recently, my coworker Hoffmeister<sup>13</sup> has described the isolation of an active substance, which he termed ecysterone, and which is more polar than ecdysone. Apparently the same substance has been isolated by Hocks and Wiechert<sup>14</sup> in the laboratory of Schering AG, Berlin. Finally, mention should be made of the moulting hormone of the crustaceans. Extracts from crustacea are active in the *Calliphora* bioassay<sup>12</sup>; this was not too surprising, since moulting of crustacea is physiologically very similar to those in insects. The moulting hormone of crayfish has been obtained in nearly pure form by Hampshire and Horn<sup>15</sup>. There are indications that all these substances are identical with our old  $\beta$ -ecdysone<sup>16</sup>.

The elucidation of the chemical structure of ecdysone was especially difficult, since only very little material was available for chemical studies. Analysis gave a composition of  $C_{7.7}H_{4.4}O_1$ , which with the actual molecular weight, gave a formula of  $C_{27}H_{44}O_6$ . Due to an error in molecular weight determination, we believed for several years in a formula  $C_{18}H_{30}O_4$ . Mainly through x-ray evidence, this was later corrected to  $C_{27}H_{44}O_6$ , a formula which requires four rings, taking into account that ecdysone is an  $\alpha,\beta$ -unsaturated ketone. A four ring structure immediately pointed to a

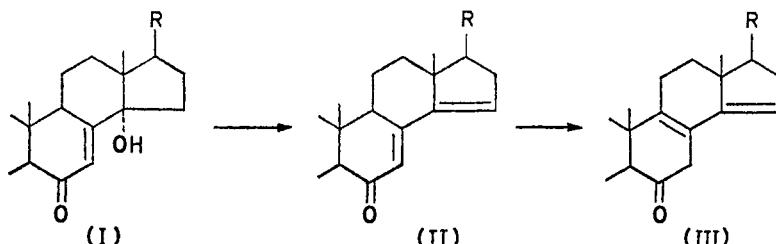
## THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

possible steroid nature; this was confirmed by a dehydrogenation experiment, which yielded methyl-cyclopentenophenanthrene<sup>17</sup>. Further information was obtained through the n.m.r. spectrum; it showed two angular methyl groups, thus confirming the steroid nature, and several hydroxyl groups, one of which must be located at C-25. The double bond had only one hydrogen, thus giving the structure



The assignment of this structure to a position in the ring system posed some difficulties<sup>18</sup>; finally, we identified it as the  $\Delta^7$ -6-ketone.

Of special importance for the structure of ecdysone was its decomposition in acid solution, yielding two substances: (i) a ketone with two double bonds in conjugation, (ii) a ketone with two conjugated double bonds no longer conjugated to the carbonyl group. The latter is the more stable product. This transformation could only be explained by a hydroxyl group in C-14, which in acid medium would be split off as water:

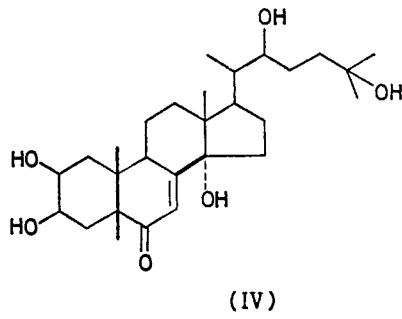


The compound (II) would rearrange to compound (III) which in this case is the more stable one, since there is less strain in ring B/C.

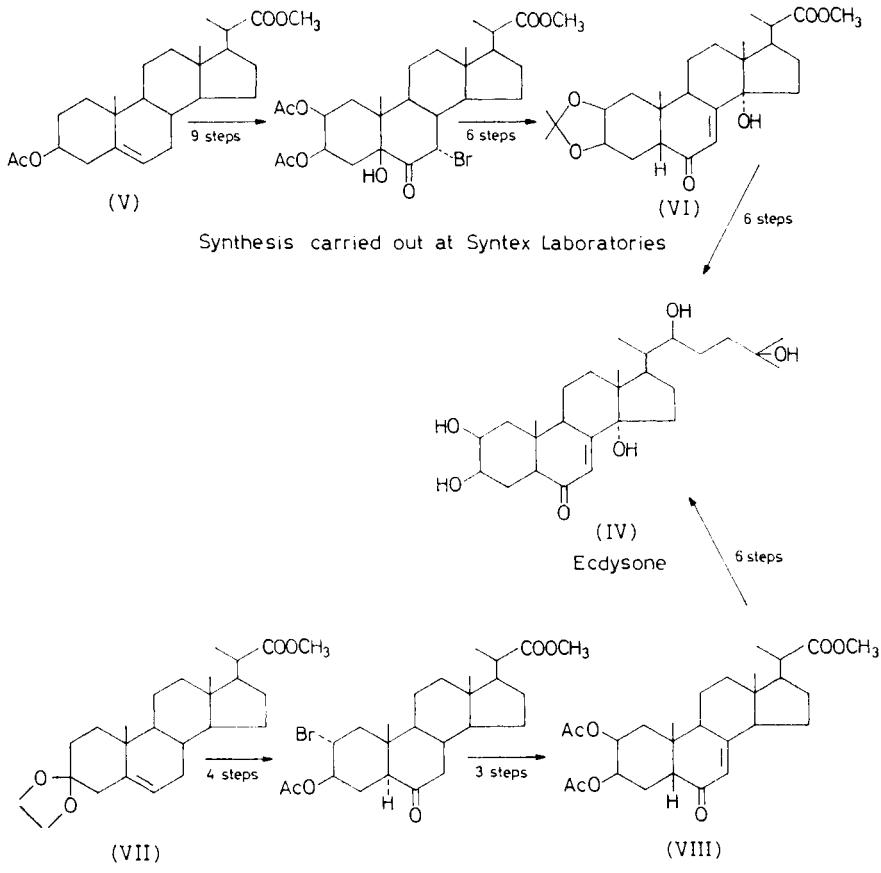
Thus, one hydroxyl group was assigned to C-14; a second one was located in the side chain, *viz.* at C-25, as was evident from the n.m.r. spectrum: there was no hydrogen beneath the two terminal methyl groups. Position 3 could be expected to bear an oxygen function, in this case hydroxyl, and since there were indications of a glycol grouping, another hydroxyl was tentatively assigned to either C-2 or C-4. For the hydroxyl unaccounted for, a position in the side chain was discussed<sup>19</sup>. The full structure, including stereochemistry, was finally elucidated by the x-ray work of Huber and Hoppe<sup>20</sup>. We had first tried to provide them with a heavy atom derivative of ecdysone; unfortunately, these derivatives did not crystallize well enough, and lack of material prevented further studies in this direction. From ecdysone crystals with the dimensions of  $0.45 \times 0.35 \times 0.15$  mm (weighing about 30 µg), Huber and Hoppe determined by x-ray-scattering 3400 structural parameters. Using a new technique<sup>20,21</sup>, they were able to calculate therefrom a complete Fourier synthesis of the ecdysone molecule, showing all carbon and oxygen atoms and about half of the hydrogens<sup>22</sup>. The structure thus determined can be described as  $2\beta,3\beta,14\alpha,22R,25$ -pentahydroxy- $\Delta^7$ -5 $\beta$ -cholest-6-one (IV): This structural formula, published a year ago, has in the meantime been confirmed by two independent syntheses in the

PETER KARLSON

laboratories of Schering/Hoffmann-La Roche<sup>23</sup> and in the Syntex labora-



tories<sup>24,25</sup>. Both groups started out from a derivative of bisnorcholenic acid (V) and (VII) respectively. The Schering-Roche group obtained by 7 steps the key intermediate,  $2\beta,3\beta$ -diacetoxy- $\Delta^7$ - $5\beta$ -bisnorcholenic acid (VIII). After reduction of the carboxyl group to the aldehyde group, the

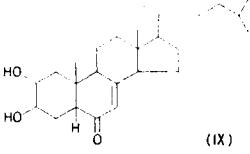
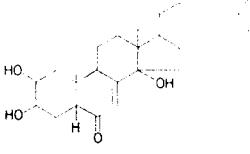
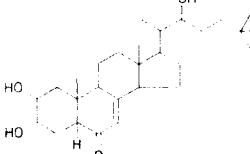
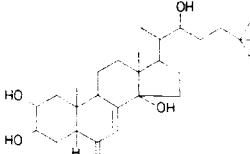
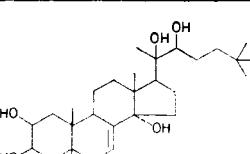


## THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

latter was reacted with a 5-carbon fragment to yield the side-chain; as the last step, the hydroxyl group at C-14 was introduced by direct oxidation with  $\text{SeO}_2$ . The syntex group introduced the  $14\alpha$ -hydroxyl earlier and obtained through 15 steps the 6-keto- $5\beta$ -cholenic acid (VI) with three hydroxyl functions; also, the side chain was introduced by a different route. Both syntheses yielded a product identical with natural ecdysone in its physical, chemical, and biological properties.

Another interesting question is the specificity of the structure in respect to the biological activity. The chemical syntheses have made available a certain number of analogues which have been assayed in the *Calliphora*

*Table 1.* Activities of the analogues of ecdysone.  
(The values given are the average of several determinations)

<i>Compound</i>	$\text{WD}_{50}$	<i>Relative activity to ecdysone</i>	<i>Modification of natural ecdysone</i>
 (IX)	4	1/80	Minus three OH-groups
 (VII)	2.5	1/50	Minus two OH-groups
 (VI)	0.75	1/15	Minus one OH-group
 (V)	0.05	1	No change
 (X)	0.04	1.25	Plus one OH-group

## PETER KARLSON

bioassay. *Table 1*<sup>26</sup> lists the compounds that are active. The last compound listed is the natural 20-hydroxyecdysone (X) isolated by Hocks and Wiechert<sup>14</sup>, presumably identical with ecdysterone<sup>18</sup>, crustecdysone, and  $\beta$ -ecdysone<sup>16</sup>. It should be mentioned that the products of the 5 $\alpha$ -series (with a *trans* junction of rings A/B) are inactive. Likewise, compounds with the  $\alpha,\beta$ -unsaturated ketone in Ring C (synthesized in our laboratory<sup>17</sup>) are inactive.

It is somewhat surprising that only two hydroxyl groups (compound IX) are necessary for biological activity. However, the number of substances assayed is so small that any generalizations seem premature.

### 3. Biochemistry of ecdysone

For a biochemist, the biochemistry of a compound, *i.e.*, the route of biosynthesis and the enzymatic degradation, is of main interest. In the case of a hormone, its biochemical mechanism of action seems to be even more important.

Studies on the biosynthesis of ecdysone have revealed that ecdysone is derived from cholesterol; the detailed biochemical pathway of its biosynthesis remains to be elucidated. This is a very difficult task, since the enzyme(s) will be found predominantly in the corresponding gland, the prothoracic gland, which in most species is a tiny, delicate structure. It seems virtually impossible to get even milligram quantities of this tissue for enzymatic studies.

However, it is not too difficult to study the precursor-product relationship with radioisotopes. We have used tritium-labelled cholesterol; this was injected into 1000 mature *Calliphora* larvae, the animals killed 36 hours later and extracted for ecdysone. Purification of this extract by solvent fractionation and paper chromatography yielded a radioactive fraction with the same R<sub>f</sub>-value as ecdysone. The eluate of this chromatogram was mixed with non-radioactive ecdysone and crystallized to constant specific activity, demonstrating that the ecdysone derived from the cholesterol-treated larvae was indeed radioactive<sup>27</sup>.

It was to be expected that ecdysone is formed from cholesterol, since insects are unable to form sterols from mevalonate or squalene; they rely on dietary sources for their needs<sup>28</sup>. Thus, cholesterol is an essential nutrient for most, if not all insects; it can be replaced to a certain extent by other plant sterols. In the beetle *Dermestes vulpinus*, up to 95 per cent of the sterol may be sitosterol; but 5 per cent has to be cholesterol. This result led Clark and Bloch to interesting speculations about steroid hormones in insects, which turned out to be correct.

As for the route of ecdysone biosynthesis, it can be expected that hydroxylations by oxygen and NADPH play a major role. It has recently been shown<sup>29</sup> that insects are capable of converting cholesterol to 7-dehydrocholesterol; this may be the first step towards the  $\Delta^7$ -6-ketone.

Turning to the metabolism of ecdysone, it is to be expected that there is an enzymatic mechanism for the inactivation, though part of the hormone is excreted with the faeces in an apparently unchanged form<sup>30</sup>. Recent experiments have shown that ecdysone is indeed degraded *in vivo* as well as *in vitro*. Further work on this subject is in progress.

## THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

The mechanism of action of ecdysone has been studied in great detail in the last three years in our laboratory, mainly in collaboration with my coworker, Dr. Sekeris. It would be a special lecture in itself, if I were to present here all the data obtained. In brief, the mechanism of action can be represented by the scheme given in *Figure 2*.

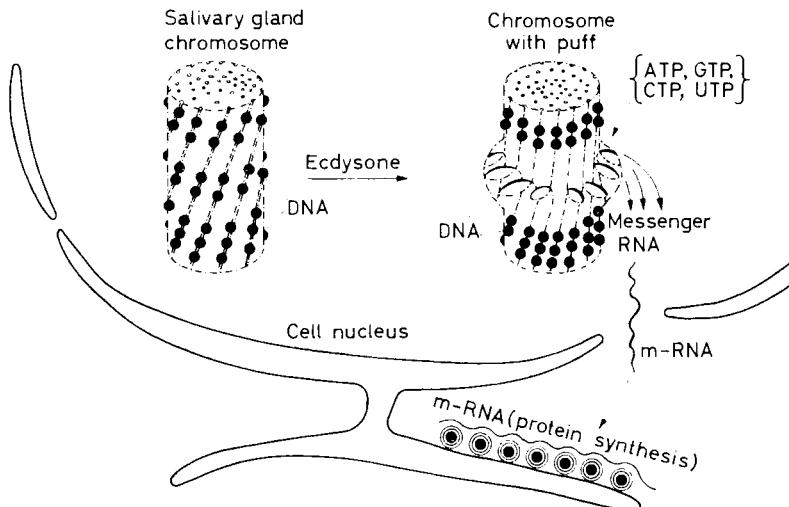


Figure 2. Mechanism of action of ecdysone

The site of action is the cell nucleus. The earliest effect observed *in vivo* is the induction of puffs in giant chromosomes<sup>31</sup>; they become visible 15 min after ecdysone injection into the whole animal. The dose needed to give a positive response is extremely small,  $2 \times 10^{-6}$  µg per animal of 20 mg weight. Puff induction means the activation of genes; it is known that the puffs are sites of active RNA synthesis, and this RNA is presumably messenger RNA, carrying the genetic information into the cytoplasm<sup>32</sup>. Increase of RNA formation can indeed be measured after ecdysone injection into *Calliphora* larvae<sup>33</sup>; this RNA has all the characteristics of messenger RNA<sup>34</sup>. Even isolated nuclei respond to ecdysone with enhanced RNA synthesis<sup>35</sup>. The messenger RNA is believed to combine with ribosomes and direct protein synthesis; in case of an enzyme protein, this process will be termed "enzyme induction". Ecdysone induces the synthesis of the enzyme dopa decarboxylase<sup>36,37</sup>. It has even been possible to demonstrate that ecdysone stimulates the production of the messenger RNA carrying the information for this enzyme. When the messenger RNA fraction from hormone-induced animals is incubated with an *in vitro* system of protein synthesis, dopa decarboxylase is formed<sup>38</sup>. Thus, all essential steps for the mechanism given above are well documented. For a detailed discussion of this work, the reader is referred to recent reviews<sup>39,40</sup>.

### 4. The juvenile hormone

As mentioned in the introduction, the juvenile hormone is produced in the corpora allata. Its biological function is the determination of larval characters at the moult, so that moulting to the next larval instar occurs.

## PETER KARLSON

Very little can be said about the chemical nature of the juvenile hormone except that it is highly controversial<sup>41</sup>. Substances with biological activity are apparently widespread in nature, they have been obtained from micro-organisms, plants, invertebrate, and vertebrate animals<sup>6</sup> and even from newspapers<sup>42</sup>. We found<sup>43</sup> that excreta from the mealworm, *Tenebrio molitor*, yield active extracts, and my former coworker Schmialek<sup>44</sup> isolated therefrom farnesol and farnesal, both being active in the bioassay. However, the activity is rather small compared with the original natural source, an extract from male *Hyalophora cecropia* moths<sup>45</sup>. Though derivatives from farnesol, like the methyl ether or the corresponding amine, are more active than the parent substance, most workers regard it as unlikely that the natural hormone is one of these compounds.

Recently, the purification of the "natural" juvenile hormone has been described by several groups. Williams and Law<sup>46</sup> isolated a crystalline material, identified as methyl-9,10-epoxy-hexadecanoate, but this carried the true hormone as an impurity, since a synthetic sample was devoid of activity. Meyer *et al.*<sup>47</sup> described a 300 000 fold purification through a number of counter-current distributions, adsorption chromatography and gas-liquid chromatography (GLC). However, the yield was very small and the active principle proved to be unstable. The molecular weight was estimated around 300. Röller *et al.*<sup>48</sup> reported a similar purification method, based mainly on thin layer chromatography and GLC, and isolated a single peak from GLC in which the activity was concentrated. The amount obtained was again very small (microgram quantities), and chemical identification was not yet possible. The molecule is considerably larger than the farnesol derivatives used as reference.

It will be of special interest to see if the juvenile hormone is also a steroid, or if it is a terpenoid, as might be implied by the activity of farnesol derivatives.

The chemical investigations of other insect hormones<sup>49,50</sup> has not grown much beyond preliminary extractions and crude preparations, so that they shall not be covered in this lecture.

## II. INSECT PHEROMONES

### 1. Terminology

The term "pheromone" has been introduced by Lüscher and myself in 1959<sup>51</sup>. It embraces chemical substances acting as messengers *between individuals* of the same species. In this respect they differ from hormones which correlate certain tissues or organs *within* the individual<sup>52,53</sup>. In a sense, pheromones create a "chemical language" for communication or rather signalization.

The classical example of a pheromone is the sex attractant of a butterfly or moth<sup>54-56</sup>. It is produced by the female in special glands at the tip of the abdomen and attracts the male moth over considerable distance. It is perceived by the antennae, *i.e.*, through the "chemical sense", and elicits a characteristic behavioural response: the male becomes excited, flutters its wings, approaches the female and finally copulates. In the field, the sex attractants are presumed to play an important role in the assembly of the sexes in efficient mating. The scent substances are carried with the wind, and

## THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

the migration of the male is directed against the wind rather than by a chemical gradient. Thus, an exact measurement of the concentration of the substance in the air is unnecessary for the animal<sup>57,58</sup>.

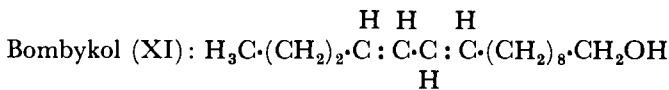
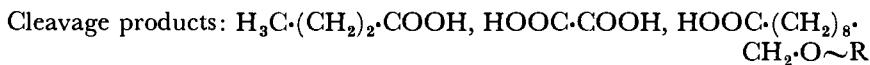
A large number of species have been shown to produce sex attractants (see the list in ref. 56). Only the few that have been isolated and characterized will be dealt with here.

### 2. Bombykol, the sex attractant of the silkworm

Pioneering chemical work on the sex attractants has been done by Butenandt and coworkers. They used the commercial silkworm, *Bombyx*, and extracted the active substance from the abdomen tips of virgin females. Several hundred thousand females were raised and processed in that way. Since only females could be used, and the males had to be sorted out to avoid copulation, it was fortunate that the work described above on the isolation of ecdysone could make use of the male pupae; thus, both programmes were carried forward with the same crop of cocoons.

The purification of the substance proved to be very difficult. An important step forward was the esterification of the pheromone (which is an alcohol) with *p*-nitroazobenzene-carboxylic acid. This derivative, a coloured substance, was easier to handle and purify. For the determination of the biological activity it had to be saponified, since the esters are inactive.

After many years of study, Butenandt's group finally succeeded in obtaining the pure ester of the *Bombyx* sex attractant<sup>59</sup>. About 12 mg were obtained in crystalline, pure form. The empirical formula of the parent alcohol termed "Bombykol" was C<sub>16</sub>H<sub>30</sub>O; two conjugated double bonds are present in the molecule. The elucidation of the structure<sup>60</sup> was done (after elaborate studies of the method with model substances) by oxidative cleavage at the double bonds with less than one milligram of the isolated ester; it can be regarded as a masterpiece of classical microchemistry. The cleavage products were butyric acid, oxalic acid, and the *p*-nitroazobenzoate of  $\omega$ -hydroxy-decanic acid. From the fragments, the structure of bombykol was easily reconstructed:



The stereochemistry is  $\Delta^{10}$ -trans- $\Delta^{12}$ -cis. The formula was confirmed by synthesis, which also made available the other three stereoisomers<sup>61</sup>. The natural isomer is by far the most active.

A few words should be said about the threshold values. During the purification, Butenandt *et al.*<sup>62</sup> used the behaviour bioassay carried out in the following manner: Male moths are kept in individual cages. A glass rod is dipped in a very dilute petroleum ether solution of the attractant and held in front of the moth. In case of a positive response, the male flutters its wings and begins a whirling dance, eventually trying to copulate with the glass rod. This assay is not very accurate, even when large numbers of animals

## PETER KARLSON

(up to 60) are used per dilution tested. Only differences in concentrations of 1:10 can be detected with certainty.

A more elaborate assay method is the recording of electroantennograms. As mentioned above, the attractant is perceived by the male through the chemoreceptors of the antennae. It is possible to insert micro-electrodes into the antennae and record the stimulation of the receptor cell; this recording is known as an "electroantennogramm"<sup>63</sup>.

From the electrical response under standardized conditions, Boeckh *et al.*<sup>64</sup> determined the threshold value necessary to stimulate single chemoreceptors. It turned out that about one hundred molecules per cell second suffice to elicit an electrical response. He also studied the behaviour reaction<sup>†</sup> in comparison with the electrophysiological data. For a positive behaviour response, only 200 bombykol molecules per cm<sup>3</sup> air are needed. In this threshold situation, 40 out of 40 000 receptor cells specialized for the sex pheromone are stimulated per second. The chemoreceptors thus function as a "molecule counting device".

### 3. Other sex attractants

Bombykol was the first sex pheromone of the lepidoptera to be isolated and identified. The choice of *Bombyx* was for practical reasons—*Bombyx* is a highly domesticated animal well suited for laboratory work. In other laboratories, species of economic importance have been predominantly studied, since sex pheromones may become important for controlling insect pests. The sex attractant of the gypsy moth, *Pophetria dispar*, has been extracted from moths collected in the field. Jacobson and Beroza reported the isolation of the pheromone in pure form and its identification as 10-acetoxy-*cis*-Δ<sup>7</sup>-hexadecanol-1<sup>65</sup>; a synthetic sample prepared by Jacobson *et al.*<sup>66</sup> was found identical in physical properties and biological activity with the natural substance, while the *trans*-isomer was not an attractant. However, the same compound, 10-acetoxy-*cis*-Δ<sup>7</sup>-hexadecanol-1, has been synthesized by other laboratories, and though the physical data were confirmed, these preparations were inactive in the behavioural as well as the electrophysiological assay<sup>67</sup>. The discrepancy has not yet been resolved.

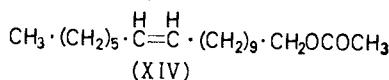
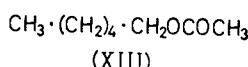
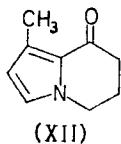
Also controversial is the isolation of the sex pheromone of *Periplaneta*, the American cockroach. The substance is produced by virgin females and can be extracted from filter-paper on which the females have been kept. Wharton *et al.*<sup>68</sup> reported the purification of this extract and the isolation of the pure attractant therefrom in microgram quantities. Jacobson *et al.*<sup>69</sup> using a different method of collection and purification, obtained 12 mg of a substance to which a cyclopropyl structure was assigned. This structure, however, proved to be incorrect<sup>70</sup>, and the chemical nature of the cockroach pheromone is still open<sup>71</sup>.

Sex attractants are not only produced by the female, but also by the male. In males of the tropical butterfly, *Lycora ceres ceres*, a peculiar pair of glands,

<sup>†</sup>According to these studies, the threshold determined by Butenandt *et al.* in the routine bioassay has to be corrected. It is still not clear for what reason the old, low values have been found in so many experiments.

## THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

the so-called hairpencils, are found. They can be protruded from the tip of the abdomen and are rich in a secretion. Extraction of this material and analysis with gas liquid chromatography yielded three fractions, which were identified<sup>72</sup> as 7-methyl-2,3-dihydropyrrolizidin-1-one (XII), cetyl acetate (XIII) and  $\Delta^{11}$  *cis*-vaccenyl acetate (XIV).



It is remarkable that two of the components bear striking similarities to bombykol and the gypsy moth attractant. Will the sex pheromones of the lepidoptera all belong to this class of straight-chain, more-or-less unsaturated aliphatic alcohols or esters? This remains to be seen. As for the heterocyclic compound, it is not clear whether it serves as part of the pheromone or rather as a defence substance.

Another male sex pheromone is found in the Indian water bug, *Belostoma indica* (= *Lethocerus indicus*). It was analysed by Butenandt and Tam<sup>73</sup> in 1957 and found to be  $\Delta^2$ -hexenyl acetate. This paper can indeed be regarded as the first chemical identification of a pheromone in insects. Besides the acetate, the butyrate is also present as a minor component in the secretion<sup>74</sup>. Though field observations on the role of this substance have not been made, there is little doubt that they act as pheromones.

Finally, mention should be made of a pheromone attracting both sexes. It can thus be classified as an "assembling scent". However, it is very likely that it facilitates mating, and that the evolutionary value of the substance is due to this fact rather than that it accounts for mass attacks of some trees by this pest.

The pheromone is produced in the hind gut of the male Bark beetle, *Ips confusus*, and is secreted with the faeces. It is only produced by animals feeding in a suitable tree. Laboratory rearing of the beetles in mass cultures provided the starting material for a chemical investigation<sup>75</sup> of the substance responsible for the attracting activity. Through solvent fractionation and gas-liquid chromatography, a substance was obtained which travelled in GLC between nonanal and geranyl acetate, and was highly attractive in the bioassay.

### 4. Pheromones of the social insects

The phenomenon of social organizations in insects has fascinated not only biologists. In our modern technical language, we can state that the community of, for instance, a bee hive or an ant's nest, must rely on a suitable system of communication between the members in order to cope with the needs. A large part of this communication uses the chemical language of pheromones<sup>76</sup>. Space does not allow me to review the large body of chemical evidence on the nature of these pheromones, many of which are related to terpenoids<sup>77</sup>.

## PETER KARLSON

Most of the substances analysed so far are derived from ants or bees. The pheromones of the termites have found only little attention. In collaboration with Prof. Lüscher, Bern, we have begun a chemical investigation of the trail pheromone of a termite, *Zootermopsis nevadensis*. A large number of termites of this species were collected in California. The animals were washed with ether and the solvent evaporated, leaving a greasy residue which is highly attractive to workers of this species. Amounts of  $\mu\text{g}$  per cm are sufficient to lay down a trail which is confidently followed in the behaviour assay<sup>78</sup>. The active principle is steam-volatile. Actually, it is rather difficult to concentrate solutions (even if the solvent is as volatile as pentane) without losses. Preliminary studies with gas-liquid chromatography shows that the substance probably has a rather small molecular weight the order of 100. However, it is premature to draw any conclusions on the chemical nature of this substance.

[*Note added in proof*.—Juvenile hormone has been identified as methyl ester of 7-ethyl-3,11-dimethyl-10-epoxy-2,6-tridecanoic acid [H. Röller, K. H. Dahm, C. C. Sweeley, and B. M. Trost. *Angew. Chem.* **79**, 190 (1967)].

## References

- <sup>1</sup> V. B. Wigglesworth. *Physiology of Insect Metamorphosis*. Cambridge University Press, London (1954).
- <sup>2</sup> P. Karlson. *Vitamins & Hormones* **14**, 227 (1956).
- <sup>3</sup> O. Pflugfelder. *Entwicklungsphysiologie der Insekten*. 2. Aufl. *Probleme der Biologie*. Bd. 5. Akad. Verlag Ges., Geest & Portig, Leipzig (1958).
- <sup>4</sup> V. J. A. Novak. *Insektenhormone*. Verlag Tschech. Akad. Wiss. Prague (1959).
- <sup>5</sup> M. Gersch. *Vergleichende Endokrinologie der wirbellosen Tiere*. Akad. Verlag Ges. Geest & Portig, Leipzig (1964).
- <sup>6</sup> M. Gabe, P. Karlson, and J. Roche. In *Comparative Biochemistry*, Vol. VI. M. Florkin and H. S. Mason (Editors). Academic Press, New York (1964).
- <sup>7</sup> P. Karlson. *Naturwissenschaften* **53**, 445 (1966).
- <sup>8</sup> E. Becker. *Biol. Zbl.* **61**, 360 (1941).
- <sup>9</sup> G. Fraenkel. *Proc. Roy. Soc. B.* **121**, 1 (1935).
- <sup>10</sup> A. Butenandt and P. Karlson. *Z. Naturf.* **9b**, 389 (1954).
- <sup>11</sup> E. Shaaya and P. Karlson. *Develop. Biol.* **11**, 424 (1965).
- <sup>12</sup> P. Karlson. *Ann. Sciences Natur.* **18**, 125 (1956).
- <sup>13</sup> H. Hoffmeister. *Angew. Chem.* **78**, 269 (1966).
- <sup>14</sup> P. Hocks and R. Wiechert. *Tetrahedron Letters* 2989 (1966).
- <sup>15</sup> F. Hampshire and D. H. S. Horn. *Chem. Commun.* **2**, 37 (1966).
- <sup>16</sup> D. H. S. Horn, E. J. Middeton, and J. A. Wunderlich. *Chem. Commun.* **2**, 339 (1966).
- <sup>17</sup> P. Karlson, H. Hoffmeister, W. Hoppe, and R. Huber. *Liebigs Ann. Chem.* **662**, 1 (1963).
- <sup>18</sup> H. Hoffmeister, C. Rufer, H. H. Keller, H. Schairer, and P. Karlson. *Chem. Ber.* **98**, 2361 (1965).
- <sup>19</sup> P. Karlson, H. Hoffmeister, H. Hummel, P. Hocks, and G. Spitteller. *Chem. Ber.* **98**, 2394 (1965).
- <sup>20</sup> R. Huber and W. Hoppe. *Chem. Ber.* **98**, 240 (1965).
- <sup>21</sup> W. Hoppe. *Acta Crystallogr. (Kbh.)* **10**, 750 (1957).
- <sup>22</sup> W. Hoppe. *Angew. Chem.* **77**, 484 (1965).
- <sup>23</sup> U. Kerb, P. Hocks, R. Wiechert, A. Furlenmeier, A. Fürst, A. Langemann, and G. Waldvogel. *Tetrahedron Letters* No. 13, 1387 (1966).
- <sup>24</sup> J. P. Sidall, A. D. Cross, and J. H. Fried. *J. Am. Chem. Soc.* **88**, 862 (1966).
- <sup>25</sup> J. P. Sidall, J. P. Marshall, A. Bomers, A. D. Cross, J. A. Edwards, and J. H. Fried. *J. Am. Chem. Soc.* **88**, 379 (1966).
- <sup>26</sup> P. Hocks, A. Jäger, U. Kerb, R. Wiechert, A. Furlenmeier, A. Fürst, A. Langemann, and G. Waldvogel. *Angew. Chem.* **78**, 680 (1966).
- <sup>27</sup> P. Karlson and H. Hoffmeister. *Z. Physiol. Chem.* **331**, 298 (1963).

## THE CHEMISTRY OF INSECT HORMONES AND INSECT PHEROMONES

- <sup>28</sup> H. L. House. In: *The Physiology of Insects. Vol. II.* M. Rockstein (Ed.), p. 769. Academic Press, New York (1965).
- <sup>29</sup> W. E. Robbins, M. J. Thompson, J. N. Kaplanis, and T. J. Shortino. *Steroids* **4**, 635 (1964).
- <sup>30</sup> H. Hoffmeister, C. Rufer, and H. Ammon. *Z. Naturf.* **20b**, 130 (1965).
- <sup>31</sup> U. Clever and P. Karlson. *Exp. Cell Res.* **20**, 623 (1960).
- <sup>32</sup> C. Pelling and C. Scholtissek. *Angew. Chem.* **76**, 881 (1964).
- <sup>33</sup> P. Karlson and G. Peters. *Gen. Comp. Endocrin.* **5**, 252 (1965).
- <sup>34</sup> C. E. Sekeris, N. Lang, and P. Karlson. *Z. Physiol. Chem.* **341**, 36 (1965).
- <sup>35</sup> C. E. Sekeris, P. P. Dukes, and W. Schmid. *Z. Physiol. Chem.* **341**, 152 (1965).
- <sup>36</sup> P. Karlson and C. E. Sekeris. *Biochim. Biophys. Acta* **63**, 489 (1962).
- <sup>37</sup> P. Karlson and C. E. Sekeris. *Arch. Biochem.* **105**, 483 (1964).
- <sup>38</sup> C. E. Sekeris and N. Lang. *Life Sci.* **3**, 625 (1964).
- <sup>39</sup> P. Karlson and C. E. Sekeris. *Recent Progr. Hormone Res.* **22**, 478 (1966).
- <sup>40</sup> P. Karlson and C. E. Sekeris. *Acta Endocrinologica*. In press.
- <sup>41</sup> V. B. Wigglesworth. *Nature* **20**, 522 (1965).
- <sup>42</sup> K. Slama and C. M. Williams. *Science*.
- <sup>43</sup> P. Karlson and P. Schmialek. *Z. Naturf.* **14b**, 12 (1959).
- <sup>44</sup> P. Schmialek. *Z. Naturf.* **16b**, 461 (1961).
- <sup>45</sup> C. M. Williams. *Nature* **178**, 212 (1956).
- <sup>46</sup> C. M. Williams and J. H. Law. *J. Insect Physiol.* **11**, 569 (1965).
- <sup>47</sup> A. S. Meyer, H. A. Schneiderman, and L. J. Gilbert. *Nature* **206**, 272 (1965).
- <sup>48</sup> H. Röller, J. S. Bjerke, and W. H. McShan. *J. Insect Physiol.* **11**, 1185 (1965).
- <sup>49</sup> M. Ichikawa and H. Ishizaki. *Nature* **198**, 308 (1963).
- <sup>50</sup> G. Fraenkel and C. Hsiao. *J. Insect Physiol.* **11**, 513 (1965).
- <sup>51</sup> P. Karlson and M. Lüscher. *Naturwissenschaften* **46**, 63 (1959); *Nature* **183**, 55 (1959).
- <sup>52</sup> P. Karlson and A. Butenandt. *Ann. Rev. Entomol.* **4**, 39 (1959).
- <sup>53</sup> P. Karlson. *Ergebn. Biol.* **XXII**, 212 (1960).
- <sup>54</sup> B. Götz. *Experientia* **7**, 408 (1951).
- <sup>55</sup> A. Butenandt. *Naturwiss. Rundschau* **8**, 457 (1955).
- <sup>56</sup> M. Jacobson. *Insect Sex Attractants*. Interscience, New York (1965).
- <sup>57</sup> J. Schwinck. *Z. vergl. Physiol.* **37**, 191 (1954).
- <sup>58</sup> H. H. Shorey. *Ann. Entomol. Soc. Am.* **57**, 371 (1964).
- <sup>59</sup> A. Butenandt, R. Beckmann, D. Stamm, and E. Hecker. *Z. Naturf.* **14b**, 283 (1959).
- <sup>60</sup> A. Butenandt, R. Beckmann, and D. Stamm. *Z. Physiol. Chem.* **324**, 84 (1961).
- <sup>61</sup> A. Butenandt, E. Hecker, M. Hopp, and W. Koch. *Liebigs Ann. Chem.* **658**, 39 (1962).
- <sup>62</sup> A. Butenandt, R. Beckmann, and E. Hecker. *Z. Physiol. Chem.* **324**, 71 (1961).
- <sup>63</sup> D. Schneider and E. Hecker. *Z. Naturf.* **11b**, 121 (1956).
- <sup>64</sup> J. Boeckh, K. E. Kaissling, and D. Schneider. *Cold Spring Harbor Symposia on Quantitative Biology. Vol. XXX*, p. 263. Sensory Receptors, Cold Spring Harbor, L.J., New York (1965).
- <sup>65</sup> M. Jacobson, M. Beroza, and W. A. Jones. *Science* **132**, 1011 (1960).
- <sup>66</sup> M. Jacobson, M. Beroza, and W. A. Jones. *J. Am. Chem. Soc.* **83**, 4819 (1961).
- <sup>67</sup> D. Schneider. *Soc. Expt. Biol. Symposium* **20**, 273 (1965). In press.
- <sup>68</sup> D. R. A. Wharton, E. D. Black, C. Merritt, M. L. Wharton, M. Bazinet, and J. T. Walsh. *Science* **137**, 1062 (1962).
- <sup>69</sup> M. Jacobson, M. Beroza, and R. T. Yamamoto. *Science* **139**, 48 (1963).
- <sup>70</sup> M. Jacobson and M. Beroza. *Science* **147**, 748 (1965).
- <sup>71</sup> D. R. A. Wharton, E. D. Black, and C. Merritt. *Science* **142**, 1257 (1963).
- <sup>72</sup> J. Meinwald, Y. C. Meinwald, J. W. Wheeler, T. Eisner, and L. P. Brower. *Science* **151**, 583 (1966).
- <sup>73</sup> A. Butenandt and N. D. Tâm. *Z. Physiol. Chem.* **308**, 277 (1957).
- <sup>74</sup> V. Devakul and H. Maarse. *Analyt. Biochem.* **7**, 269 (1964).
- <sup>75</sup> D. L. Wood, L. E. Browne, R. M. Silverstein, and J. O. Rodin. *J. Insect Physiol.* **12**, 523 (1966).
- <sup>76</sup> E. O. Wilson. *Science* **149**, 1064 (1965).
- <sup>77</sup> G. W. K. Cavill and P. L. Robertson. *Science* **149**, 1337 (1965).
- <sup>78</sup> M. Lüscher and B. Müller. *Naturwissenschaften* **47**, 503 (1960).

# THE GINKGOLIDES†

Koji NAKANISHI

Department of Chemistry, Tohoku University, Sendai, Japan

## I. INTRODUCTION

The ginkgo tree (*Ginkgo biloba*; Icho in Japanese) occupies a unique position in botany for several reasons. It is the sole representative of its family and is not linked to any other living plant; its origin extends as far back as the Paleozoic. The order *Ginkgoales* was once widely distributed throughout the world but in the last few million years all species excepting *Ginkgo biloba* have become extinct, the other species being found only as fossils in petrified woods; moreover, *Ginkgo biloba* itself is believed to have remained unchanged for the last million years or so. The ginkgo tree is thus called a "living fossil" or "fossil tree". The ginkgo tree was unknown outside the Orient before the 18th century but is now rather commonly distributed in Europe, America and other continents.

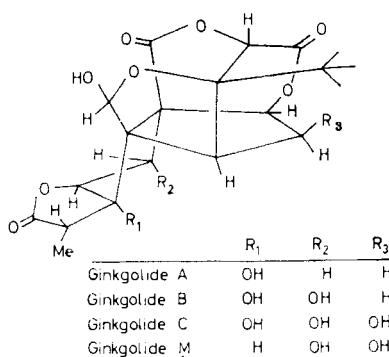
The ginkgo tree has already been subject to numerous chemical investigations, which have led to the characterization of the following compounds: the phenolic compounds, ginkgol<sup>1,2</sup>, bilobol<sup>3</sup> and ginkgolic acid<sup>1</sup> from the fruit; cyanogenetic glycosides and amino acids from the seeds<sup>4,5</sup>; the aliphatic compounds, ginnol<sup>6</sup>, ginnon<sup>7</sup>, and *n*-hexenal<sup>8</sup>, and shikimic acid<sup>9</sup> from the leaves; the bisflavone ginkgetin<sup>10-12</sup> from the leaves; and finally D-sesamin<sup>13</sup> and the sesquiterpenoid bilobanone<sup>14</sup> from the heartwood.

Study of the root itself was commenced in this laboratory in 1960 under the late Professor S. Fujise, but the early stage of the work was seriously hindered by purification problems and polymorphism.

† Shortly after the Symposium, it was discovered that dehydroginkgolide A (Figure 13) very easily underwent a photochemical rearrangement and that several properties of the rearrangement product had in fact been wrongly attributed to dehydroginkgolide A. In particular, the lack of a Cotton effect in a sample of dehydroginkgolide A (actually photodehydroginkgolide A which had been formed during storage) indicated the absence of an  $\alpha$ -hydroxylactone function in the original ginkgolides, and thus led us wrongly to believe that a hemiacetal function was present in these substances.

Once the occurrence of this photochemical rearrangement had been recognized, the existence of an  $\alpha$ -hydroxylactone in the ginkgolides became apparent; moreover, the ease with which this rearrangement occurred provided important evidence in the derivation of the current structure of the molecule.

The present content has been modified from the lecture given in Stockholm, since an article based on a wrong structure would not be of much significance. (September 1966).



Structure of Ginkgolides proposed at Stockholm (July 1966).

## KOJI NAKANISHI

For structural studies of cage molecules such as the ginkgolides, molecular models are indispensable to clarify the spatial relation of the various groups. This was especially so in the present case because the partial structures (described later in the article) were connected by intervening quaternary carbon atoms, and because the only carbocyclic ring turned out to be a spiro[4.4]nonane system, which became apparent only in the final stages of the structure elucidation. Anisotropic effects displayed in the n.m.r. spectra, especially the difference in chemical shifts when one group is changed to another (*e.g.*, oxidation of hydroxyl to ketone), the nuclear Overhauser effect, the photochemical rearrangement, and other spatial interactions, as studied with molecular models, were of extreme importance in determining the structure. However, since such an argument would be difficult to present without the extensive use of models, an attempt has been made to derive the structure as much as possible without recourse to spatial considerations. The derivation presented in the following therefore differs slightly from that actually used in the structure determination but on the whole follows the same logical sequence.

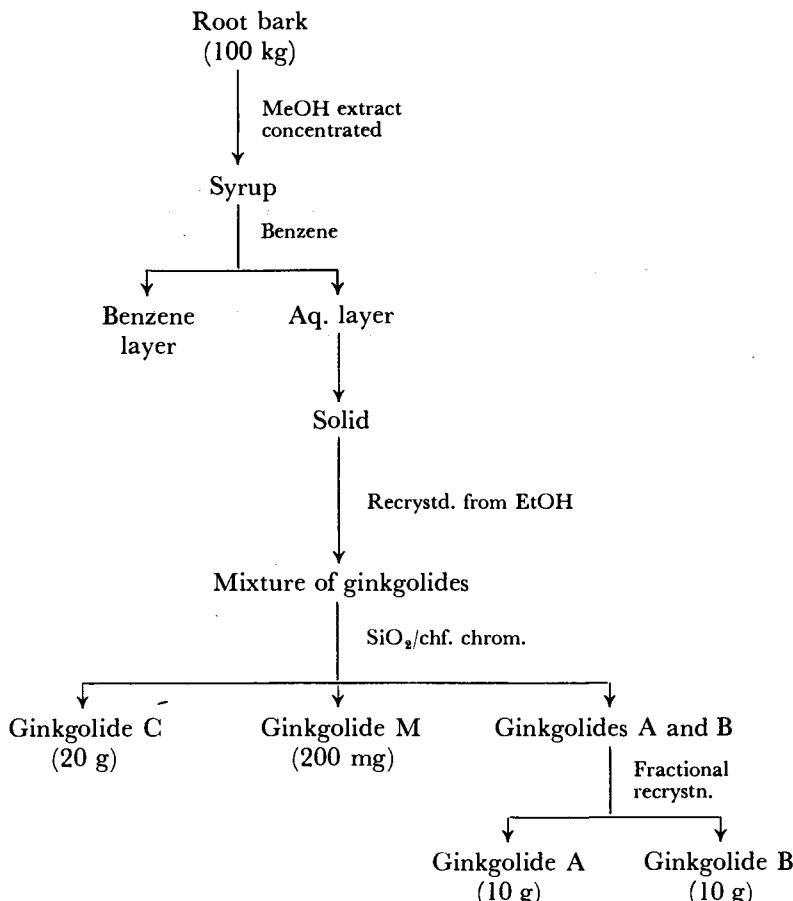
Several ginkgolide derivatives, with and without heavy atoms, had been sent to Professors M. Kakudo and Y. Sasada, Osaka University, and Professor W. N. Lipscomb and Dr. J. W. Moncrief, Harvard University, for x-ray analysis, but the crystals sent have not been easily amenable to this powerful technique, probably owing to our inappropriate recrystallization techniques or deterioration of crystals in the post. Nevertheless, we are grateful to these workers for their interest and preliminary examinations of the ginkgolides. On the other hand, it has been particularly gratifying to us to be informed, after completion of our studies, that the entire structure including the absolute configuration had been confirmed by a totally independent x-ray study (private communication from Dr. N. Sakabe, Nagoya University).

The structure elucidation based entirely on chemical reactions and spectroscopic methods has led to the disclosure of many fascinating chemical and physico-chemical phenomena, and has contributed greatly to increase our appreciation of the behaviour of complex organic molecules.

## II. ISOLATION OF GINKGOLIDES

The procedure being followed at present is outlined in *Figure 1*. Four compounds designated ginkgolides A, B, C, and M (for minor) have been isolated by this procedure. A methanol extract of the undried chopped root bark was concentrated to a syrup and extracted with benzene. The solid material which precipitated from the aqueous layer crystallized from ethanol as mixed crystals of the ginkgolides. A solution of these crystals in acetone was absorbed on Celite, the Celite was placed as a layer on top of a silica gel column, and the column was eluted with chloroform containing traces of acetone and ethanol. This afforded a ginkgolide A-ginkgolide B mixture, a small amount of ginkgolide M and finally ginkgolide C. The separation of ginkgolides A and B was extremely tedious, and satisfactory results were achieved only after a 10-15 step fractional crystallization procedure, the purity being checked either by n.m.r. or optical rotation. Furthermore,

### THE GINKGOLIDES



*Figure 1.* Scheme for the isolation of ginkgolides from *Ginkgo biloba* L. (*Ginkgoaceae*) (five trees of 30 cm in diameter were used)

separation was complicated by the strong tendency of ginkgolide A and especially ginkgolide C to exhibit polymorphism.

### III. STRUCTURE OF THE GINKGOLIDES

#### (a) Molecular formulae and general properties

The ginkgolides are  $C_{20}$  compounds but owing to the easy fragmentation of their derivatives upon electron impact, the molecular formulae could be established only with the aid of high resolution mass spectrometry using a direct inlet system. Ginkgolide A dimethyl ether gave a clean molecular ion peak at 436.168 (calc. for  $C_{22}H_{28}O_6$ , 436.173). All the ginkgolides are bitter in taste, and are extremely stable towards mineral acid such as conc.  $HNO_3$ , conc.  $HCl$ , warm conc.  $H_2SO_4$ , as well as to warm 1N  $NaOH$ . Evaporation to dryness of a conc. nitric acid solution of the ginkgolides results in recovery

KOJI NAKANISHI

of crystalline starting material. The molecular formulae and specific rotation of the ginkgolides are given in *Table 1*.

*Table 1.* Molecular formulae and specific rotation of ginkgolides

Compound	Mol. formula	Sp. rotation [ $\alpha$ ] <sub>D</sub>
Ginkgolide A dimethyl ether	C <sub>22</sub> H <sub>28</sub> O <sub>9</sub> †	—
Ginkgolide A	C <sub>20</sub> H <sub>24</sub> O <sub>9</sub>	-39° (c, 1·0, dioxane)
Ginkgolide B	C <sub>20</sub> H <sub>24</sub> O <sub>10</sub>	-63°
Ginkgolide C	C <sub>20</sub> H <sub>24</sub> O <sub>11</sub>	-19°
Ginkgolide M	C <sub>20</sub> H <sub>24</sub> O <sub>10</sub>	-39°

† Found: mol. wt. 436·168. C<sub>22</sub>H<sub>28</sub>O<sub>9</sub> requires mol. wt. 436·173.

**(b) Functional groups**

The absence of any ketone grouping in the ginkgolides is supported by the plain negative optical rotatory dispersion (o.r.d.) curves exhibiting no Cotton effect in the range 250–700 m $\mu$ . The number and nature of the hydroxyl groups were deduced by the conventional n.m.r. techniques of comparing the spectra of the original ginkgolides with those of the acetates and by measurements of n.m.r. spectra in DMSO-d<sub>6</sub> and addition of D<sub>2</sub>O or deuterated acid. The results are summed up in *Table 2*. Because of solubility problems, the n.m.r. spectra were in general measured as trifluoroacetic acid solutions: in general, the proton chemical shifts (apart from hydroxyl protons) were within ca.  $\pm$  0·2 p.p.m. of their positions when measurements were made, when possible, in deuteriochloroform. Unless stated otherwise, the chemical shifts refer to trifluoroacetic acid solutions.

*Table 2.* Studies on the functional groups present in ginkgolides

**KETONE:** None, plain optical rotatory dispersion

**HYDROXYL:** From acetate and DMSO-d<sub>6</sub> n.m.r. spectra

	2°	3°
Ginkgolide A	one	one
Ginkgolide B	two	one
Ginkgolide C	three	one
Ginkgolide M	three	—

**t-BUTYL:**

N.m.r. Spectra : 9 H(s) at 1·2–1·3 p.p.m. in trifluoroacetic acid

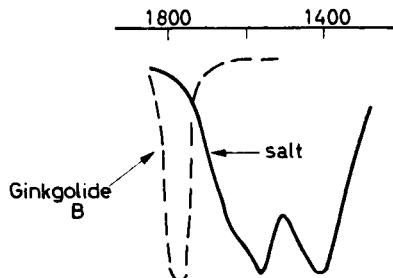
Mass Spectra : Peak at 57-074 (50–100%). Calc. mol. wt. for C<sub>4</sub>H<sub>9</sub>, 57-070  
M-57 peak in ginkgolide A-LiAlH<sub>4</sub> reduction product  
(triether) at 309 (31%)

Kuhn-Roth oxidation  
of Ginkgolide C: t-BuCOOH (*p*-bromo phenacyl ester)

The n.m.r. spectra of all derivatives, excepting photodehydro-ginkgolide A, showed a 9-proton singlet in the range 1·2–1·3 p.p.m. Since it is most

## THE GINKGOLIDES

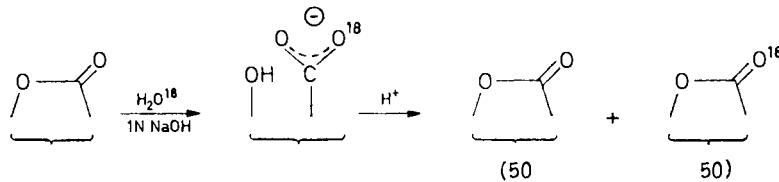
unlikely that three methyl groups attached to different carbon atoms would constantly be subject to identical overall anisotropic and electronic effects, this conspicuous signal suggested the presence of a *tert*-butyl group, for which no precedence has been recorded in the natural products field. Moreover,



I.r. spectra of ginkgolide B (KBr disk) and that of its sodium salt (in KBr/NaOH disk)

the mass spectra had a strong peak (relative intensity 50–100 per cent of base peak) at 57·074, which corresponded to the stable *tert*-butyl cation, 57·070. This unique characteristic of the ginkgolides, *i.e.*, presence of the *tert*-butyl group, was finally established by isolation of pivalic acid, characterized as the crystalline *p*-bromophenacyl ester, by oxidation of ginkgolide C under Kuhn-Roth conditions.

The i.r. spectra (KBr, acetonitrile, dioxane) of the ginkgolides show a strong but ill-defined absorption around 1790 cm⁻¹. This band was replaced by carboxylate absorptions when, for example, an aqueous sodium hydroxide



$H_2O$  containing 30%  $O^{18}$

$M^+$  peaks of ginkgolide A dimethylether (mol. wt. 436)

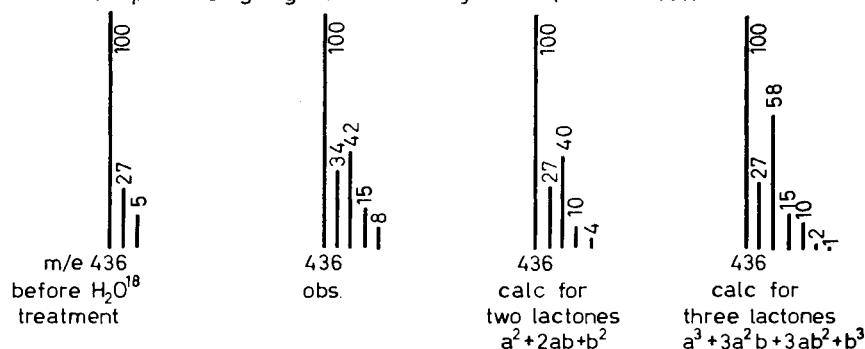


Figure 2. Attempts to establish the number of lactone rings from the mass spectrum of the  $H_2O^{18}$ -treated compound

KOJI NAKANISHI

solution of ginkgolide B was evaporated to dryness at room temperature under reduced pressure, and the residual mixture of the sodium salt and sodium hydroxide was made up into KBr disks; ginkgolide B was recovered upon dissolution of the KBr disk and acidification, thus indicating no chemical change excepting cleavage and recyclization of the lactone rings. This indicated that all three lactones are *essentially strain-free*, and further that the ginkgolides contained no ketonic carbonyl group.

The exact number of lactones had remained a problem for a long period. The first clear experimental evidence for the presence of three lactones was secured through the detailed n.m.r. analysis (cf. para. IIIk) of "ginkgolide A triether" which showed that three carbonyl groups had been reduced to methylene groups. Accordingly, a modified technique for lactone titration was developed (*Table 3*) and the results fully confirmed the conclusions drawn from the n.m.r. analysis. Attempts to derive the number of lactones by mass spectrometric measurements of ginkgolide A dimethyl ether utilizing  $\text{H}_2\text{O}^{18}$  (*Figure 2*) led to anomalous results, which on the other hand, were of use in clarifying the chemical behaviour of the lactone groups.

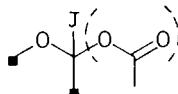
The n.m.r. spectra of the ginkgolides and derivatives were characterized by two lowfield singlets, I and J (cf. para VI) generally appearing at *ca.* 5.4 and 6.2 p.p.m. (trifluoroacetic acid), respectively. The two signals can be easily differentiated because signal J was invariably taller than signal I. The broader singlet I is due to the carbonyl proton of the secondary hydroxyl, which together with one of the lactones, constitutes an  $\alpha$ -hydroxylactone,  $-\text{CH}_1(\text{OH})-\text{CO}-\text{O}-$  (cf. para IIIg).

Lactones: three

i.r.  $1780 \text{ cm}^{-1}$  (MeCN, dioxane)  
n.m.r. of ginkgolide A-triether

Potentiometric titration  
Mass spectrometry of  
ginkgolide A-dimethylether  
( $\text{O}^{16}$  and  $\text{O}^{18}$ )

Ether



J at 6.17 p.p.m. (sharp s,  $\text{CF}_3\text{COOH}$ )  
in ginkgolide A

( ■ indicates carbon with no proton)

As regards the other lowfield singlet J, which appears as a singlet even in  $\text{DMSO-d}_6$ , we shall only mention at this stage that it is flanked by an ethereal oxygen and one of the lactone groups as shown above. The nine oxygen atoms in ginkgolide A are thus accounted for by three lactones, two hydroxyls, and one ether group.

## THE GINKGOLIDES

### (c) Number of lactone rings

A large number of potentiometric lactone titrations were carried out since in spite of the fact that conventional titrations (Method A) seemed to indicate that only two lactones were present, n.m.r. analyses of the triether and absence of other functions to account for the two remaining oxygen atoms suggested there should be three lactones. The objective of establishing the presence of three lactone rings was finally achieved by Method B in which the alkali solutions were evaporated to dryness; the i.r. spectra of the residual mixture of sodium hydroxide and the sample (KBr disk) showed that all carbonyl functions had been cleaved to carboxylate groups, and moreover no rearrangement had occurred by this treatment because unchanged starting materials were recovered upon acidification of the residues. Apparently, one lactone ring remains unopened under conditions of Method A.

*Table 3.* Potentiometric titration

(One lactone grouping is more hindered than the others. All pKs are below 6)

Compound	No. of Lactones	No. of moles of alkali consumed	
		Method A*	Method B†
I Ginkgolide A	3	2.1	2.4
II Ginkgolide B	3	1.9	2.5
III Ginkgolide C	3	1.9	2.8
IV Ginkgolide C O-methylether §	3	1.8	2.5
V Ginkgolide A O-methylether acid §	2	3.1‡	—
VI Bisnor-ginkgolide A	2	1.8	—

\* Compound in 0.1 N aq. NaOH titrated with 1.0 N HCl.

† Solution of compound in 1.0 N aq. NaOH evaporated under reduced pressure at 50°C. Residue redissolved in water (to 0.1 N NaOH) and titrated with 1.0 N HCl.

‡ One mole of alkali consumed by the free carboxyl.

§ The OMe groupings do not correspond.

Compound IV in *Table 3* is a ginkgolide C methyl ether, in which the secondary hydroxyl group of the hydroxylactone moiety is methylated; compound V is a derivative having a cleaved ring C but will not be discussed in this article; compound VI is discussed later (paras III i and l).

#### *Attempts to establish the number of lactone rings from the mass spectrum of the H<sub>2</sub>O<sup>18</sup>-treated compound*

Treatment of ginkgolide A with methyl iodide and potassium carbonate in acetone yielded the dimethyl ether, which unlike other ginkgolide derivatives, gave a clear molecular ion peak at m/e 436 (see *Table 1*).

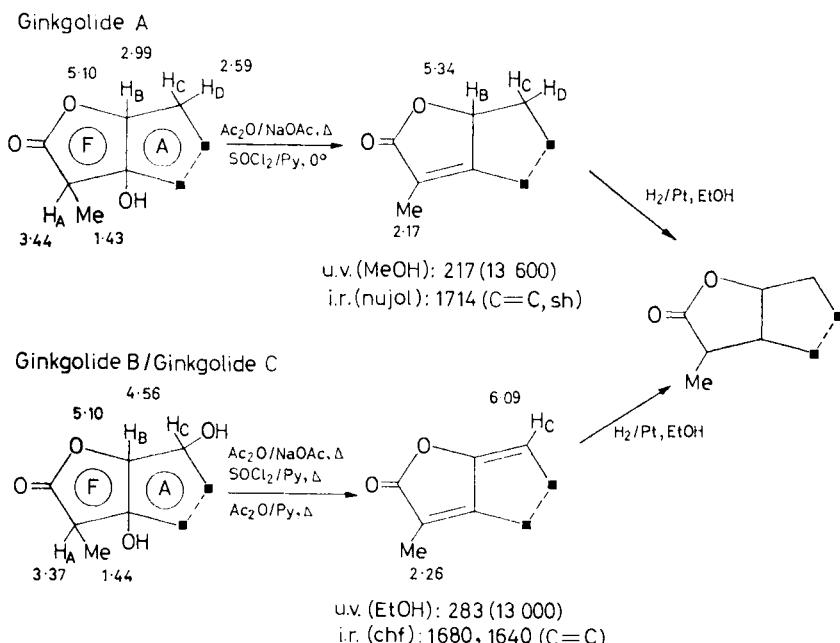
Since all the lactone groups are apparently cleaved under the conditions used in the potentiometric titrations (Method B, *Table 3*), it was thought that alkaline cleavage of the lactones, with 1N sodium hydroxide in water containing 30 per cent H<sub>2</sub>O<sup>18</sup> and under the same conditions as above, followed by recyclization on acidification would result in incorporation of O<sup>18</sup> into all the lactone groups. Assuming free rotation about the C—C axis of the carboxylate groups, and hence a 50:50 chance of O<sup>18</sup> incorporation, it should be possible to calculate the number of lactone groups in the molecule

from the ratios of the isotope peak-heights. Equations relating the intensities of the  $M + 2$ ,  $M + 4$ ,  $M + 6$ , . . . peaks and the number of lactone rings have been derived and are shown in *Figure 2*; (a) and (b) denote, respectively, the content of  $O^{18}$  and  $O^{16}$  (in the present case 0.30 and 0.70). In the calculated mass spectra shown in *Figure 2*, relative peak intensities derived from these equations are superimposed on the relative peak intensities arising from  $C^{13}$ .

Surprisingly, a comparison between the observed and calculated spectra seemed to indicate that only two and not three lactone functions were present. In view of the results of potentiometric titration, this anomaly should be interpreted by assuming an  $S_N2$  type *O*-alkyl fission of one of the lactone rings, which is followed by another inversion at the hydroxyl-bearing carbon atom upon recyclization; the  $O^{18}$  atom thus is incorporated in the opened hydroxy-carboxylate form but is expelled again upon lactone formation. Such *O*-alkyl fission in one of the lactone rings is not unexpected if one considers the severely hindered cage structure of the ginkgolides.

#### (d) Reactions of rings F/A; nature of protons A–D

Reactions and spectroscopic data leading to the partial structure comprising rings F and A are summarized in *Figures 3 and 4* (for n.m.r. see para VI). The *tertiary* hydroxyl group of ginkgolide A was readily dehydrated by



*Figure 3.* Reactions of rings F/A; nature of protons A–D. Small squares (■ .... ■) denote carbon atoms carrying no hydrogens as derived from n.m.r. data (measured in tetrafluoroacetic acid)

## THE GINKGOLIDES

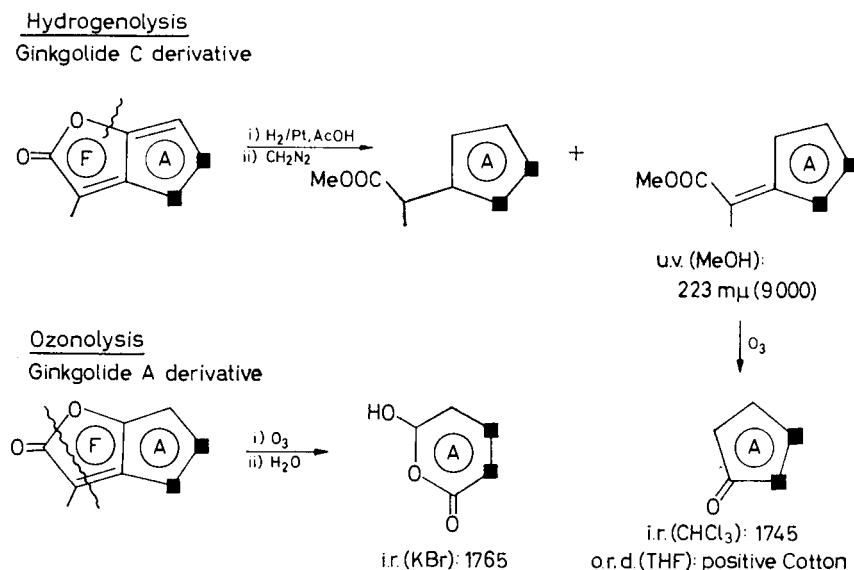
boiling in acetic anhydride-sodium acetate or treatment at 0° with thionyl chloride in pyridine, with concomitant conversion of a secondary methyl to an olefinic methyl. Furthermore, monoanhydro-ginkgolide A was readily hydrogenated to its dihydro derivative. The dehydration presumably proceeds through the acetate or thionyl ester of the tertiary hydroxyl function, since mild treatment of ginkgolide A diacetate with sodium acetate readily gave rise to monoanhydro ginkgolide A monoacetate.

Proton D is replaced by a hydroxyl function in ginkgolides B and C as indicated by n.m.r. data and formation of the dianhydro derivatives, the spectroscopic data of which are fully consistent with the  $\alpha, \beta, \gamma, \delta$ -unsaturated  $\gamma$ -lactone moiety, and with this the part structures shown are established.

The only difference between ginkgolides A and B is the presence of this additional hydroxyl group in the latter since hydrogenation of dianhydro-ginkgolide B and monoanhydro-ginkgolide A afforded the same product; ginkgolide C has one more hydroxyl group than ginkgolide B.

### (e) Cleavage of rings F and A

When the dianhydro-ginkgolide C derivative, dianhydromonomethoxy-ginkgolide C monoacetate was hydrogenated in acetic acid with platinum, hydrogenolysis occurred to afford products having the partial structures shown in *Figure 4*.



*Figure 4.* Cleavage of rings F and A under conditions of hydrogenolysis and ozonolysis

Ring A is five-membered because ozonolysis of the  $\alpha, \beta$ -unsaturated ester gave a ketone with an i.r. absorption at 1745 cm<sup>-1</sup> (CHCl<sub>3</sub>). Ozonolysis of monoanhydro-ginkgolide A monoacetate resulted in a Baeyer-Villiger type

oxidation of the product to give the lactol shown. Both reactions provide further support for the structures of rings F and A.

### (f) Protons E, F, G and H (Ring B)

The n.m.r. signals due to the moiety comprising the isolated four-proton system E–H in ginkgolide A and ginkgolide B could not be fully analysed in a straightforward manner because of the overlap of signals E–G; however, proton G is fortunately substituted by a hydroxyl in ginkgolide C and accordingly ginkgolide C and its derivatives were quite suited for n.m.r. (*Figure 5*). The spectra can only be satisfactorily explained by the arrange-

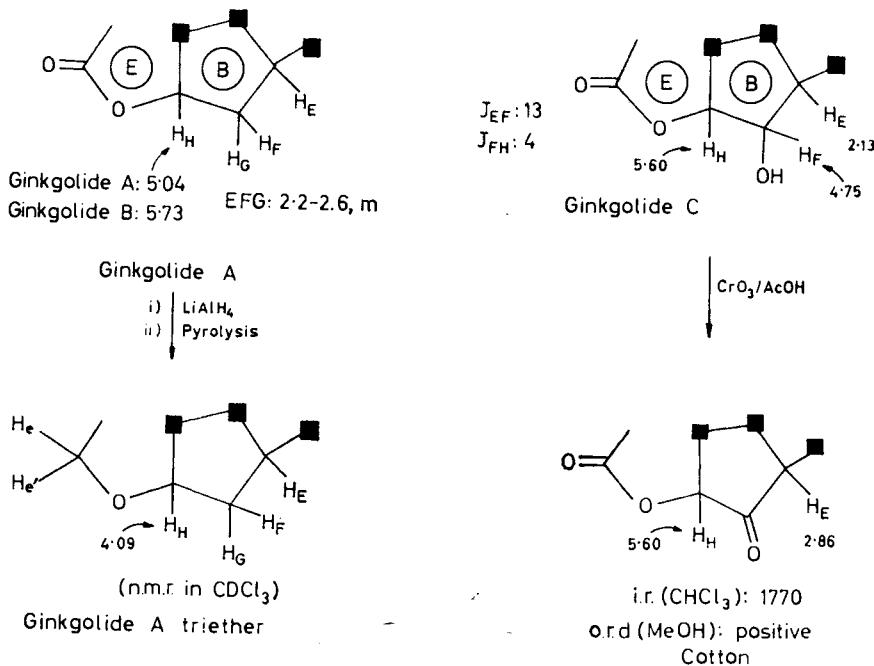


Figure 5. N.m.r. signals due to Protons E, F, G, and H (Ring B)

ment  $\text{CH}_\text{H}-\text{CH}_\text{F}\text{OH}-\text{CH}_\text{E}$ ; moreover, chromic acid oxidation of ginkgolide C yielded a ketone, the formation of which was supported by the appearance of a positive Cotton effect in the o.r.d. curve. Furthermore, this ketone group absorbed at  $1770 \text{ cm}^{-1}$  in the i.r., and accordingly it is a five-membered ring ketone.

The lowfield position of the proton H signal (5.04–5.73 p.p.m.) in the four ginkgolides obviously requires this proton to be  $\alpha$  to a strongly electron-attracting group. In the so-called ginkgolide A-triether (*Figures 6–8*), in which all lactonic carbonyl groups have been reduced to methylene groups, without any change in the ring system, the proton H signal coupled only to protons F and G and not to one of the newly introduced methylene protons. This indicates that in the original ginkgolides proton H is attached to a carbon atom linked to an oxygen, and cannot be  $\alpha$  to the carbonyl end of a

## THE GINKGOLIDES

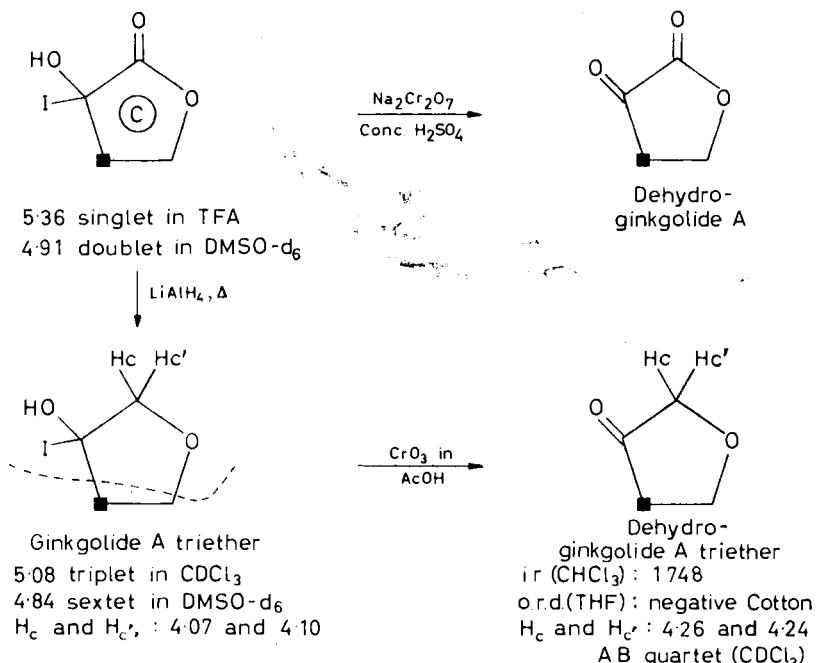


Figure 6. Proton I and the hydroxylactone (Ring C)

lactone. Moreover, in the triether, the chemical shift of proton H is shifted upfield to the extent of 0.6 p.p.m. (*i.e.*, 4.09 p.p.m. in  $\text{CDCl}_3$ ; proton H in ginkgolide A dimethyl ether, a ginkgolide A derivative which is soluble in  $\text{CDCl}_3$ , appears at 4.64 p.p.m.). Therefore, the above mentioned oxygen atom should be further linked to a carbonyl group, so that proton H is  $\alpha$  to the oxygen end of a lactone as shown in Figures 6 and 7.

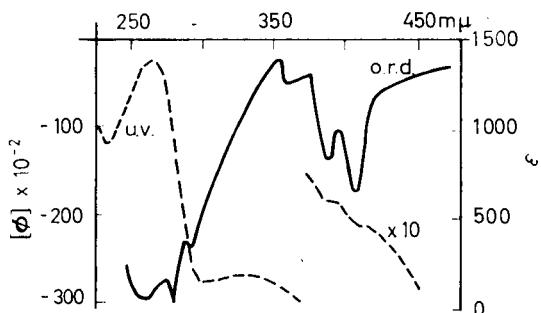


Figure 7. Optical rotatory dispersion and ultraviolet curves of dehydro-ginkgolide A

### (g) Proton I and the hydroxylactone (Ring C)

Oxidation of ginkgolide A with a reagent stronger than the usual chromic acid washing solution gave dehydro-ginkgolide A lacking the proton I; these were the only conditions under which ginkgolide A could be successfully

oxidized, but the yield was almost quantitative. The u.v. and o.r.d. curves clearly indicated that an  $\alpha$ -dicarbonyl function had been formed; thus ginkgolide A contains an  $\alpha$ -hydroxylactone group which is oxidized to a ketolactone. The presence of the hydroxylactone group received further support from the fact that (i) the n.m.r. of ginkgolide A-triether (*Figures 6 and 11*) showed a triplet (proton I) at 5.08 p.p.m. in  $\text{CDCl}_3$ , which became a sextet in  $\text{DMSO-d}_6$ ; (ii) a clear M-60 peak was observed in the mass spectrum. Finally, the hydroxylactone is considered to be five-membered because oxidation of the triether gave a ketone (dehydro-ginkgolide A triether, see *Figure 6*) with an i.r. band at  $1748\text{ cm}^{-1}$ .

### (h) The three lactones

Partial structures incorporating lactones F, E, and C, have been derived in the foregoing paragraphs, but since the points of attachment of the carbonyl group in lactone E and the oxygen in lactone C have not been defined these two lactone groupings could be one and the same. This point can be clarified from the following observation. When ginkgolide A was heated for 30 min at  $160^\circ$  in 50 per cent sodium hydroxide and then acidified, a product designated bisnor-ginkgolide A, because of the loss of two carbon atoms ( $\text{C}_{18}\text{H}_{24}\text{O}_7$ , from mass spectrum), was obtained in *ca.* 50 per cent yield. Bisnor-ginkgolide A has only *two* lactone rings (by titration, *Table 3*), and it was clear that lactones F and E had not undergone any change since the n.m.r. signals arising from protons A, B-D, and H-E remained essentially the same (excepting that proton E is further split by *ca.* 9 c/s, see *Figure 9*). Instead of the hydroxylactone function, bisnor-ginkgolide A contains a *hemiacetal* as evidenced by its oxidation to a trilactone. Convincing evidence that it is the hydroxylactone C which has been lost is provided by the detection of oxalic acid in the alkali reaction mixture, *i.e.*, the two carbon

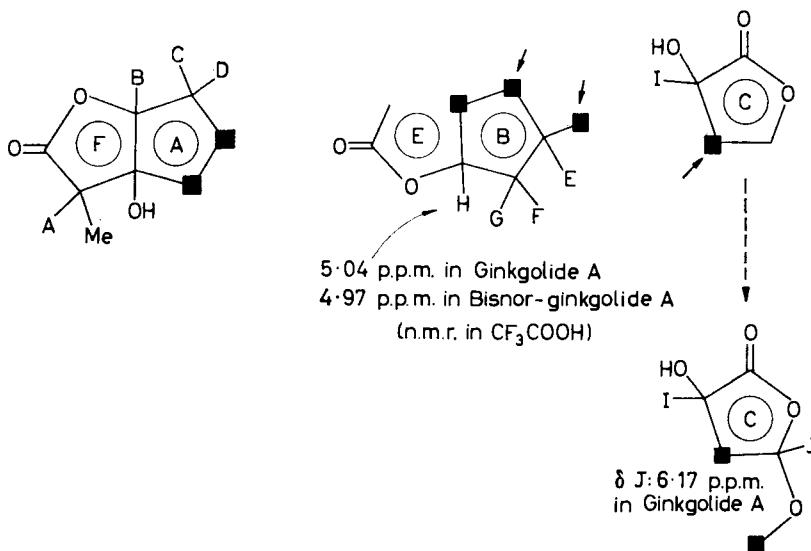


Figure 8. Nature of the three lactones

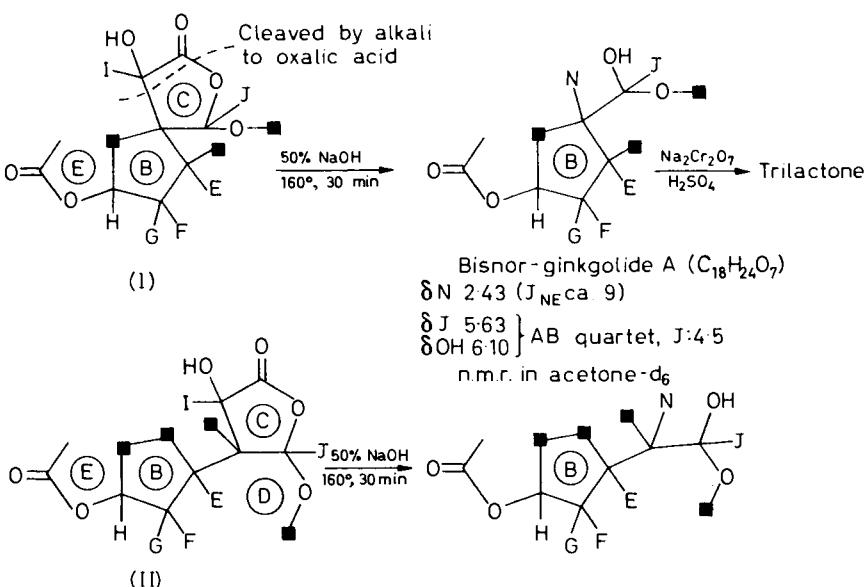
## THE GINKGOLIDES

atoms lost in ginkgolide A originate from the hydroxylactone group, which therefore *cannot* be identical with lactone E. The three lactone groups C, E, and F are therefore as shown in *Figure 8* and account for eight of the nine oxygen atoms in ginkgolide A.

As already mentioned (in para III b) the n.m.r. spectra of the ginkgolides exhibit a conspicuous *singlet* J at the lowfield of *ca.* 6.2 p.p.m., besides that of the other lowfield proton I on ring C. The appearance of the proton J signal at such lowfield can only be accounted for by its presence in one of the partial structures  $-\text{O}-\text{CH}_J-\text{CO}-\text{O}-$  or  $-\text{O}-\text{CH}_J-\text{O}-$ . Since the n.m.r. spectrum of the triether (*Figures 6 and 11*) establishes that in ginkgolide A *one* of the lactone carbonyl groups has no  $\alpha$ -proton and that protons A and I are  $\alpha$  to the other two carbonyls, respectively, it follows that proton J must be located between the ether oxygen and the oxygen of lactone C, viz.,  $-\text{C}(\text{OH})(\text{H}_I)-\text{CO}-\text{O}-\text{C}(\text{H}_J)-\text{O}-\text{C}-$ .

### (i) The formation of bis-norginkgolide A

As mentioned in *Figure 8*, vigorous alkali treatment of ginkgolide A afforded bisnor-ginkgolide A, resulting from the loss of a two carbon chain presumably in the form of glyoxylic acid (which yields the oxalic acid mentioned previously via a Cannizzaro reaction). The n.m.r. spectrum indicated that the only changes involved were related to ring C, and that rings A, B, E, and F remained intact. A hemiacetal function is present in bisnor-ginkgolide A since: (i) the presence of a carbonyl proton J is indicated by the lowfield signal at 5.63 p.p.m., which forms part of an AB type quartet



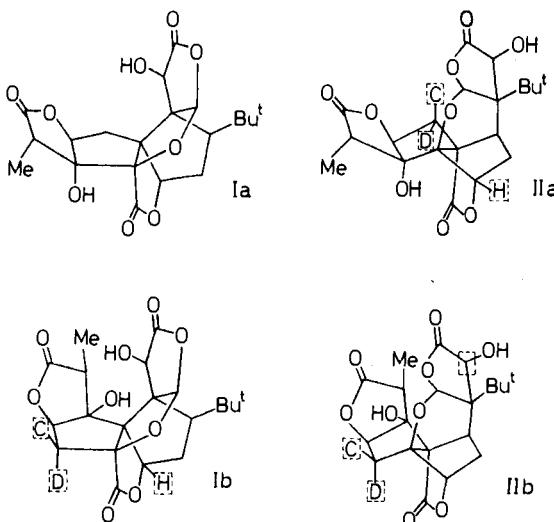
*Figure 9.* The formation of bis-norginkgolide A

with a hydroxyl signal, in acetone-d<sub>6</sub>; and (ii) oxidation gave a trilactone consisting of lactones E, F, and the new lactone derived from the hemiacetal.

The bisnor compound was extremely informative in that it was the only derivative that had a proton (N) strongly coupled to the hitherto isolated proton system E-H through proton E, the coupling constant J<sub>NE</sub> being *ca.* 9 c/s. Undoubtedly, this new proton N should be placed at the blocked carbon in ring C (see short straight arrows in *Figure 8*) after loss of the two-carbon fragment, and the large N-E coupling further requires protons N and E to be on adjacent carbon atoms, *i.e.*, *ring C should be attached to one of the blocked carbons (marked with arrows) adjacent to proton E*, and this leads to the two possibilities (I) and (II) depicted in *Figure 9*. In further confirmation of these two partial structures, proton N is slightly coupled to proton J.

#### (j) The possible structures

We are now left with the problem of linking rings F/A and the *tert*-butyl group with either part structure (I) or (II). Only four pairs of structures (*Figure 10*) can be constructed, each pair differing only in the mode of attachment of rings F and A (cf., Ia and Ib). Of these four pairs, *only two* (pairs I and II) are in accord with the minimum requirement for a facile photochemical rearrangement (cf. para III m), namely close proximity of the *tert*-butyl group and the hydroxyl function of ring C.



*Figure 10. The possible structures*

The structures (Ib) and (IIa) are eliminated on the following grounds. The proton H signal is at 5.04 p.p.m. in ginkgolide A, but this is subject to considerable deshielding when proton C or D is replaced by a hydroxyl group, *i.e.*, it appears at 5.73 p.p.m. in ginkgolide B, 5.60 p.p.m. in ginkgolide C, and 5.49 p.p.m. in ginkgolide M. The protons in question are too

## THE GINKGOLIDES

widely separated in structures (Ib) and (IIa) to account for this spatial interaction.

Structure (IIb) is also eliminated by comparing the chemical shifts of protons C and D in ginkgolide A-triether (all carbonyl groups reduced to methylenes, no skeletal change) and dehydro-ginkgolide A-triether (secondary hydroxyl in the triether oxidized to a carbonyl, see *Figure 6*). Thus in the latter compound, protons C and D have undergone an upfield shift to the extent of 0.34 and 0.61 p.p.m., respectively, but the structure (IIb) cannot account for this large anisotropic effect of the carbonyl group.

We are now left with only one structure, (Ia), which is fully substantiated by all experimental data, both chemical and spectroscopic, some of which are discussed below. For the sake of simplicity, structures in the following figures depict the relative and absolute stereochemistry as well.

### (k) Ginkgolide A-triether and the deuterotriether

Lithium aluminium hydride reduction of ginkgolide A in dioxane, acid decomposition of the complex, and heating of the resulting polyol at 150°/5 mmHg for 2–3 hours, afforded a key derivative which we named ginkgolide A-triether (*Figures 5 and 6*). The M<sup>+</sup> peak indicated that in terms of molecular weight, the only change was the conversion of all three lactonic carbonyl groups into methylene groups, and in fact, full analysis of its n.m.r. spectrum and comparison with the ginkgolide A spectrum clearly showed that no skeletal rearrangement was induced by this reaction sequence.

This “triether” was of great importance since it permitted characterization of the moieties connected to the lactone groups. The n.m.r. spectra of the ginkgolides are too simple because they consist of several unrelated (*i.e.*, uncoupled) proton-systems A–D, E–H, I and J, isolated by fully substituted carbon atoms or lactone rings. In the triether, however, the too-simple n.m.r. spectrum of ginkgolide A is converted into a more complex spectrum, thus enabling one to derive more information on the arrangements of protons. Although the ginkgolide A triether spectrum was quite complex, the entire region was analysed in detail by taking full advantage of solvent shifts to reveal overlapping signals, and using the techniques of double and triple resonance. The spectrum measured in deutero-chloroform after addition of heavy water is shown in *Figure 11*; the more simple methyl resonance region is omitted. The protons introduced by the lithium aluminium hydride reduction step are indicated in lower case lettering in the *Figure*. All chemical shifts and coupling constants shown in the *Figure* were deduced by first-order analysis, and several small long-range couplings were also disclosed.

After full analysis of this complex spectrum the n.m.r. of the deuterotriether, in which all protons expressed in lower case lettering have been replaced by deuterium, was measured and it was gratifying to find the n.m.r. assignments shown in the upper trace to be fully confirmed. A comparison of the two traces shows that all signals and large couplings involving protons indicated in lower case lettering have disappeared in the deutero derivative. Moreover, the close similarity between the spectrum of the deutero

compound and that of ginkgolide A (*Figure 15*) should be noted. This is of course not unexpected if one considers that the difference between the two compounds, resides only in the replacement of the three CO groups with CD<sub>2</sub> groups. Thus the overall picture in the ginkgolide A to triether to

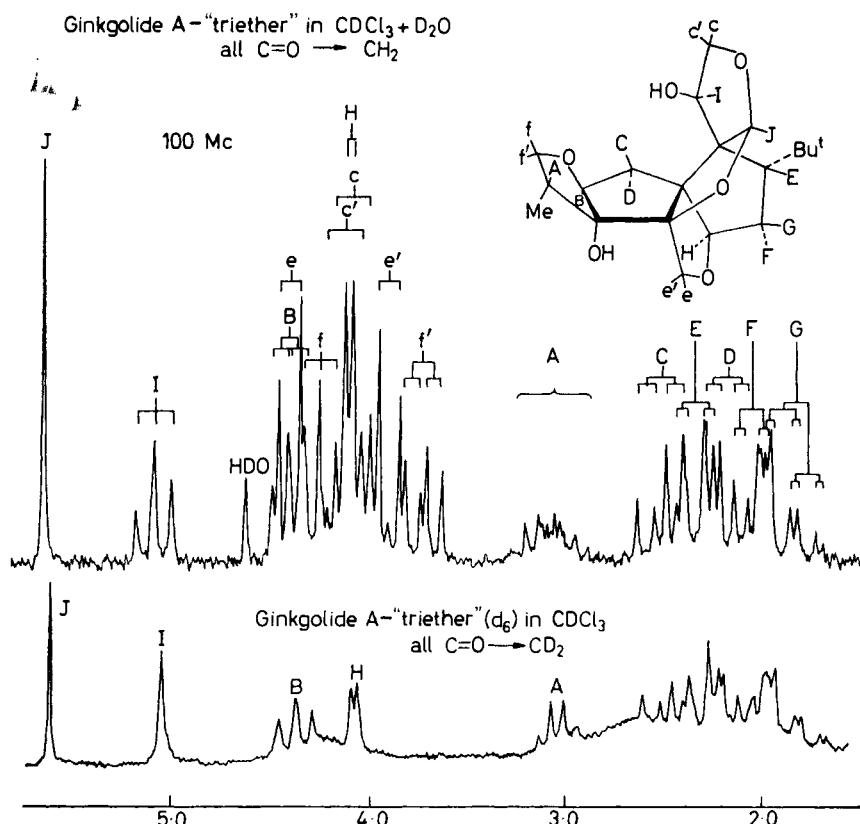


Figure 11. N.m.r. spectra of Ginkgolide A-triether and the deuterotriether

deuterotriether conversions is, complication of the spectrum by changing CO to CH<sub>2</sub>, and resimplification by changing CH<sub>2</sub> to CD<sub>2</sub>.

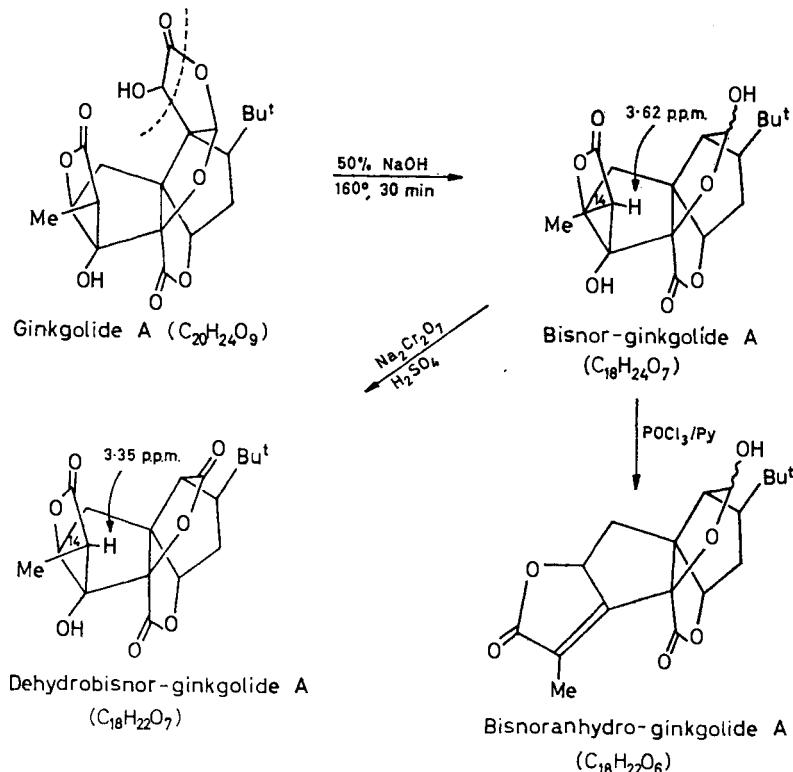
Finally, the spectrum of the ketone shown in *Figure 6*, *i.e.*, the dehydro-triether, was also fully analysed, and a comparison with the original triether spectrum revealed significant shifts, especially in signals C and D (cf. para III j), resulting from the anisotropy of the newly introduced ketone group. All other significant shifts were fully consistent with the derived structure.

### (1) Bisnor-ginkgolide A

Bisnor-ginkgolide A and its dehydro derivative, also played important roles in clarifying the surroundings of the lactone groups (*Figures 8 and 9*),

### THE GINKGOLIDES

and can be represented as shown in *Figure 12*. Removal of the two-carbon chain in the form of glyoxylic acid (detected as oxalic acid) is presumably due to alkaline cleavage of the 1,3-diol formed upon opening of lactone ring C.



*Figure 12.* Reactions of bisnor-ginkgolide A and related compounds

#### (m) Photodehydroginkgolide A

It was mentioned in *Figure 6* that drastic oxidation of ginkgolide A afforded dehydro-ginkgolide A containing the  $\alpha$ -ketolactone ring. However, examination of the properties of this straightforward oxidation product was complicated by the ready occurrence of an unsuspected photochemical rearrangement.

For a long time, the secondary hydroxyl was considered to form part of a hemiacetal group since the o.r.d. curve of the oxidation product was thought to exhibit no Cotton effect above  $250 \text{ m}\mu$ , a behaviour in accord with a "tetralactone" structure. However, it was later discovered that the o.r.d. curve measured actually belonged to the photodehydro-ginkgolide A which had been formed during storage of the dehydro-ginkgolide A (*Figure 13*). Only after measuring the n.m.r. spectrum of a sample that had been kept for some time did it become apparent that the dehydro-ginkgolide A had been

completely converted into photodehydro-ginkgolide A and that the *tert*-butyl group had surprisingly been converted into a gem-dimethyl group and a methylene group.

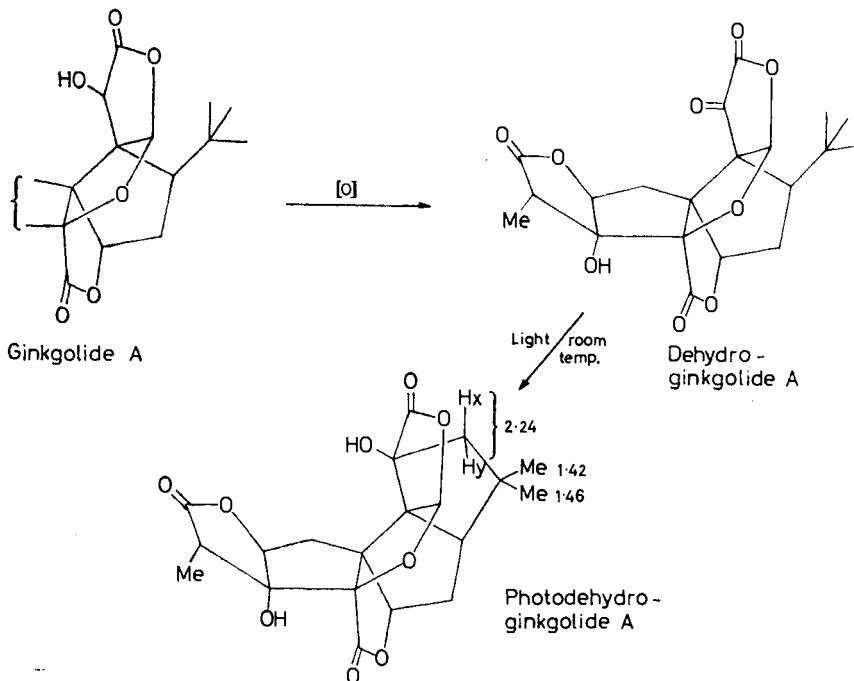


Figure 13. Formation of Photodehydroginkgolide A

This suggested that a photochemical reaction had taken place. The u.v., o.r.d. (Figure 7) and n.m.r. spectroscopic data on the fresh oxidation product indeed revealed the presence of a keto-lactone system and the intact *tert*-butyl group. The dehydro-ginkgolide A was rapidly transformed into the photo-rearrangement product, photodehydro-ginkgolide A, by either leaving a solution or a spot on a thin-layer chromatographic plate in the light; in the latter case the conversion was complete after 4 hours. Apparently, this exceptionally facile photo-cyclization is caused by the long-wavelength u.v. absorption band (Figure 7).

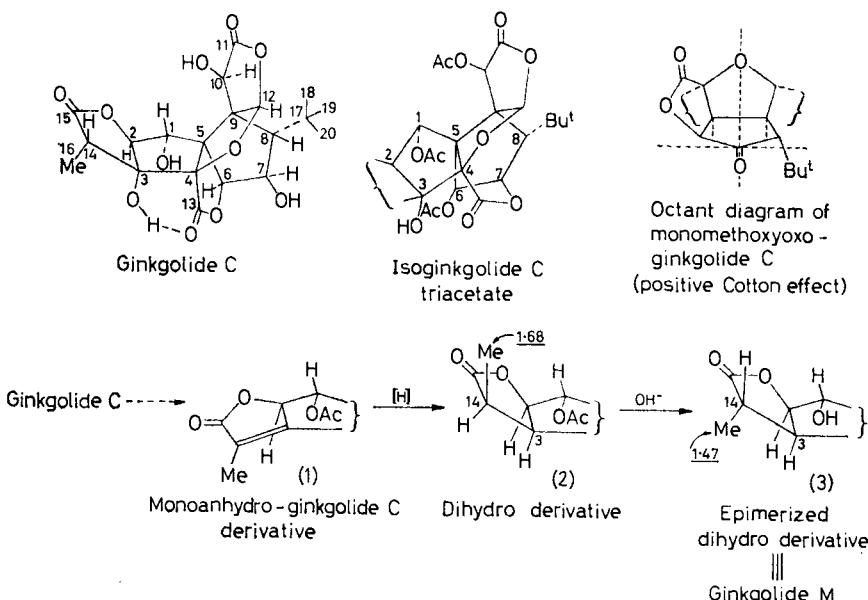
A model of ginkgolide A indicates that the spatial disposition of the *tert*-butyl group ( $\alpha$ -configuration) and the carbonyl group is ideally suited for hydrogen abstraction from one of the *tert*-butyl methyl groups to give photodehydro-ginkgolide A, which now has a unique cage structure built up of seven strain-free five-membered rings.

#### IV. STEREOCHEMISTRY OF THE GINKGOLIDES

The configurations at seven asymmetric centres, C-1, C-2, C-3, C-7, C-8, C-10, and C-14 have to be considered for ginkgolide C, which contains the largest number of asymmetric carbon atoms.

## THE GINKGOLIDES

The *absolute stereochemistry* of the ginkgolides can be simply deduced as follows. The facile photochemical rearrangement of dehydro-ginkgolide C requires the *tert*-butyl group and the C<sub>9</sub>–C<sub>10</sub> bond to be situated on the *same side* of ring B. Since monomethoxyoxo-ginkgolide C (OMe at C–10 and ketone at C–7) exhibits a positive Cotton effect in its o.r.d. curve, it follows from the octant diagram (*Figure 14*) that the ginkgolides must have the absolute stereochemistry depicted in *Figure 14*. The remaining configurations were established as follows.



*Figure 14.* Stereochemistry of the ginkgolides

(i) C–10: Proton I must be on the same side of ring C as the *tert*-butyl since these groupings participate in an intramolecular Overhauser effect (see Table under *Figure 15*).

(ii) C–7: When ginkgolide C is acetylated by boiling in Ac<sub>2</sub>O–NaOAc, an isotriacetate is readily formed which is reconverted to ginkgolide C itself upon saponification with MeOH–NaOMe. The isotriacetate can be further acetylated to isoginkgolide C tetraacetate. These acetates are isomeric to the normal tri- and tetraacetates, and are formed by translactonization of lactone E (from C–6 to C–7, see *Figure 15*). This requires an S-configuration at C–7 which is in line with the coupling constants J<sub>EF</sub> 13 c/s and J<sub>FH</sub> 4 c/s in ginkgolide C, and J<sub>EF</sub> 0 c/s and J<sub>FH</sub> 4 c/s in the isomeric acetates. The detected nuclear Overhauser effects in the normal and isomeric acetates is also consistent with the structures shown.

(iii) Rings F/A: Cleavage of ring F with alkali and acidification readily gives back the starting material, thereby indicating that the  $\gamma$ -lactone is *cis*-fused.

## KOJI NAKANISHI

(iv) *C-14 and C-3*: A *trans* arrangement of 14-H and 3-OH is suggested by the ready formation of monoanhydro-ginkgolide C under mild conditions (Figure 3).

This is established by the following evidence which is described in only very brief form. A monoanhydro-ginkgolide A derivative 1 was hydrogenated to the corresponding dihydro derivative 2, which upon mild base treatment was epimerized at C-14 to the dihydro derivative 3, which was identical in every respect with ginkgolide M. The C-14 epimerization was evident from the appearance of the methyl signals of 2 and 3, at 1.68 and 1.47 p.p.m., respectively. Since hydrogenation must have occurred from the same side of ring F in compound 1, the first dihydro derivative 2 has the 14-H and 3-H in a *cis* relation, while the epimerized dihydro derivative, which is ginkgolide M, has them in a *trans* arrangement and is also the thermodynamically more stable. Furthermore, the ring F/A configurations in all ginkgolides are identical as indicated by comparisons of the chemical shifts of protons H<sub>B</sub>, H<sub>C</sub> and J<sub>BC</sub>, and the fact that mild base treatment of the natural products induces no epimerization of the 14-Me group. Namely, the 14-methyl group adopts the thermodynamically more stable configuration.

(v) *C-3 and C-2*: The 3-OH and 2-H are  $\alpha,\alpha$ -*cis*. This conclusion is based on the fact that the chemical shift of 14-H, which as deduced above is *trans* to the 3-OH, is considerably affected by the oxygen function at C-12 and hence should be  $\beta$ . Namely, it is at 3.62 p.p.m. in bisnor-ginkgolide A (Figure 12) but at 3.35 p.p.m. in dehydrobisnor-ginkgolide A, the 0.27 p.p.m. diamagnetic shift being in line with the spatial arrangement of the C-12 oxygen function and a 14- $\beta$ H.

The ready methylation of the 3-OH in ginkgolide A upon treatment with MeI/K<sub>2</sub>CO<sub>3</sub> in acetone can also be understood on the basis of an intra-molecular H-bonding between the 3-OH and 13-CO (i.r. evidence), which would serve to weaken the O-H bond and facilitate methylation.

(vi) *C-1*: The 1-OH in ginkgolides B, C, and M adopts an  $\alpha$ -configuration because the 6-H is subject to considerable deshielding by introduction of the additional OH. Molecular models show that a 1 $\alpha$ -OH is in the sample plane as that of 6-H. The 6-H chemical shifts are as follows: ginkgolide A, 5.04; ginkgolide B, 5.73; ginkgolide C, 5.60; ginkgolide M, 5.49 p.p.m.

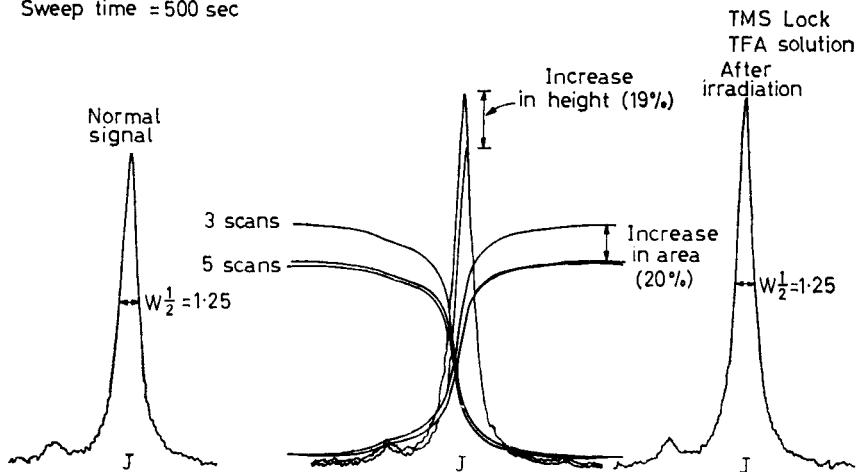
## V. THE NUCLEAR OVERHAUSER EFFECT (NOE)

In an attempt to relate the isolated proton systems, by using n.m.d.r. (double resonance) techniques to search for some indication of small long range couplings between protons belonging to different systems, it was found that irradiation (saturation) of the *tert*-butyl group caused a significant increase (*ca.* 30 per cent) in the height of the I and J proton signals, but did not cause any great decrease in the half-band width of these signals. Moreover, in the case of ginkgolide C, the heights of signals due to protons E and F were also found to increase on irradiation of the *tert*-butyl group. Since it seemed unlikely that all four of these protons could be coupled to the *tert*-butyl by a through-bond mechanism, the possibility of through-space coupling or an effect involving a relaxation mechanism was considered.

# THE GINKGOLIDES

Proton J at 6.26 p.p.m.

Sweep width = 50 c/s  
 Sweep time = 500 sec



## NUCLEAR OVERHAUSER EFFECTS\*

<i>Proton observed</i>	<i>Proton irradiated</i>	Ginkgolide A	Ginkgolide B	Ginkgolide C	Ginkgolide C tetraacetate	Isoginkgolide C tetraacetate	Isoginkgolide C tricarboxylate	Ginkgolide A diOMe, monoCO <sub>2</sub> Me
I	<i>t</i> -Butyl	30 [32]	25 [24]	20 [24]	33 [36]	27 [37]	30 [32]	Nil [<4]
J	"	22 [22]	21 [23]	20 [19]	22 [21]	Nil	Nil	22 [28]
F	"	-	-	4 [10]	16 [13]	19 [22]	24 [23]	-
E	"	-	-	-	6 [16]	13 [17]	14 [14]	-
H	"	Nil	Nil	Nil	Nil	6 [10]	12 [15]	Nil
J	E	-	-	Nil	Nil	10 [17]	7 [21]	-
E	J	-	-	Nil	Nil	23 [30]	20 [27]	-

\* Given as the percent increase in integrated intensity on irradiation; the figure in parenthesis is the percent increase in height of the signal. Solutions were in trifluoroacetic acid; signals recorded at 50 c/s/500 sec sweeps and integrated at 50 c/s/100 sec sweeps, on a Varian HA-100.

Figure 15. The nuclear Overhauser effect: Effect of irradiating the *tert*-butyl at 1.30 p.p.m. ginkgolide C

In a series of recent communications<sup>15</sup>, it has been shown that in certain rigid molecules, it is possible to observe an intramolecular Overhauser effect when a large contribution to the relaxation of a particular proton comes from one or more closely situated protons in the same molecule. In such a case, saturation (*e.g.*, by double resonance) of the proton(s) responsible for the relaxation of the proton in question causes an appreciable increase (up to

KOJI NAKANISHI

50 per cent) in the integrated intensity of the signal due to the latter proton<sup>15</sup>.

The results of chemical and spectroscopic studies carried out indicate that the ginkgolides possessed a rigid cage-like skeleton, and it therefore appeared very likely that the increase in the height of the signals from the I, J, E, and F protons in ginkgolide C, on irradiation of the *tert*-butyl group, might well arise from a nuclear Overhauser effect.

Integration of the I, J, E, and F proton signals both before and during irradiation of the *tert*-butyl group showed that a marked increase in the integrated intensities of these signals resulted from saturation of the *tert*-butyl group. Some results are listed in the *Table* (under *Figure 15*) and an example is shown in *Figure 15*. On the other hand, irradiation of proton J, for example, did not cause any appreciable increase in the *tert*-butyl signal. In the case of isoginkgolide-triacetate and -tetraacetate (see *Figures 14 and 15*), isomers of the normal acetates, it is of interest to note that proton J does not

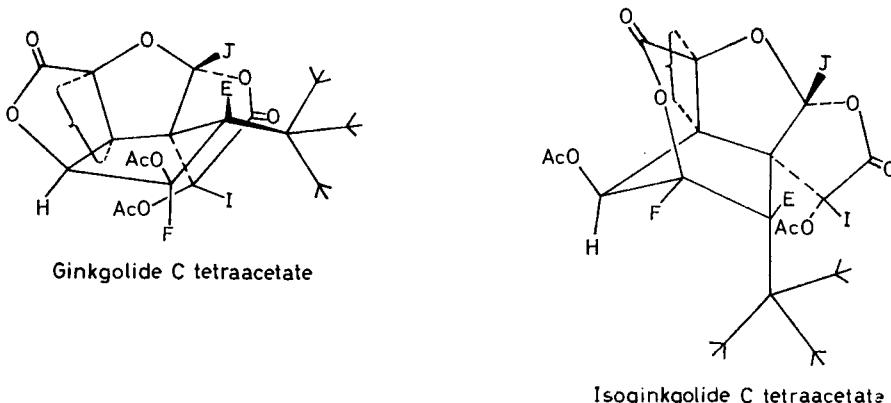
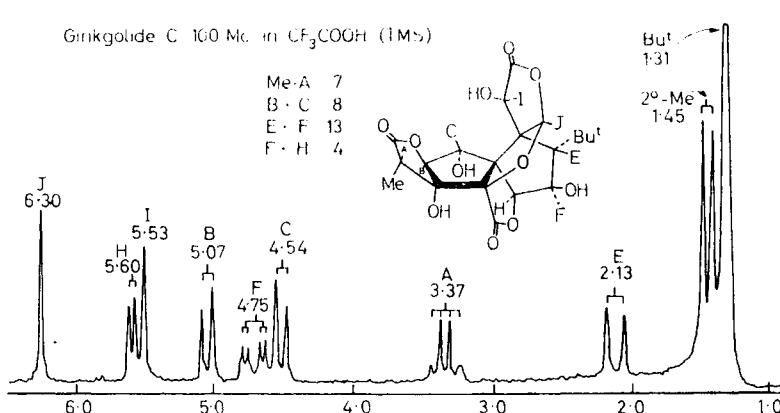
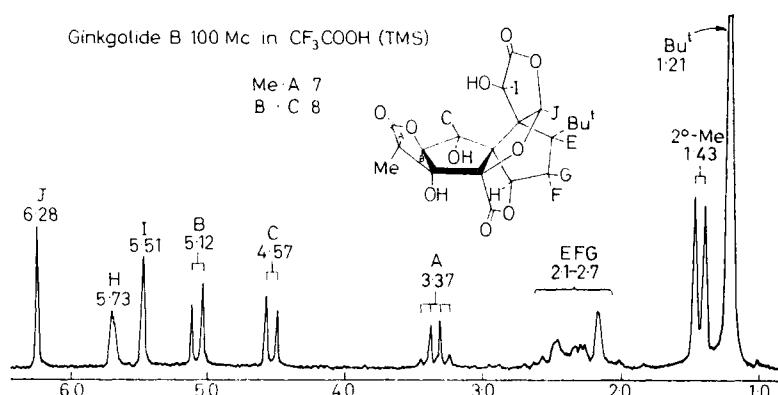
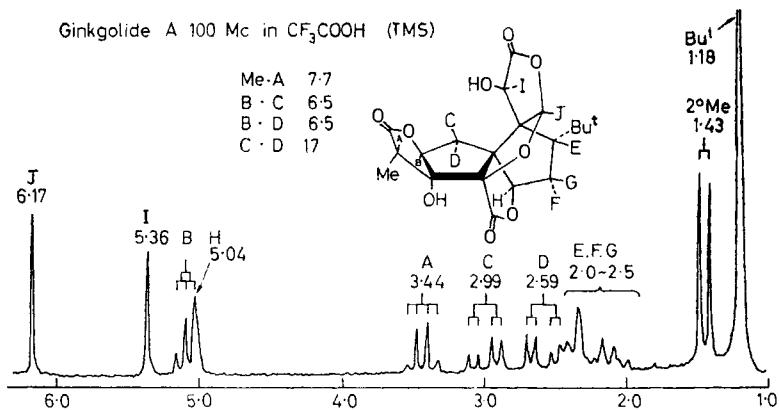


exhibit any nuclear Overhauser effect but instead the *tert*-butyl protons now appear to be largely responsible for the relaxation of proton H which is not so affected in the case of the other compounds in the *Table*. It was significant that the nuclear Overhauser effect was not exhibited any more by proton I in the ginkgolide A derivative shown in the last column of the *Table* (in this derivative, lactone C is opened and the resulting three OH groups at 10, 12 and 13 are methylated). The importance of these results in assigning structures and stereochemistry to the ginkgolides is obvious, since in order for this relaxation mechanism to operate the protons in question must be very closely situated<sup>15</sup>.

## VI. N.M.R. SPECTRA OF GINKGOLIDES A, B and C

The relatively simple n.m.r. spectra of these complicated compounds (*Figures 16–18*) need little comment and will not be discussed further. However, it may be commented that due to the rigid and extensive cage

### THE GINKGOLIDES

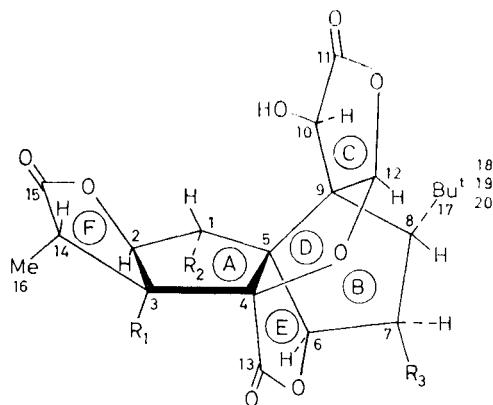


*Figures 16–18. N.m.r. spectra of ginkgolides A, B and C*

structure, it should not be surprising if the presence of bond-angle distortions and bond anisotropy effects in these molecules lead to rather unusual chemical shifts and coupling constants in their n.m.r. spectra. For this reason, in most of the described n.m.r. analyses, only the differences in chemical shifts between two or more derivatives were considered. Moreover, the familiar relation between dihedral angle and coupling constant has not been unduly employed.

### VIII. FULL STRUCTURE OF THE FOUR GINKGOLIDES

The ginkgolide structure is built up very comfortably from six five-membered rings (*Figure 19*), and the chemical inertness of the ginkgolides and their rigid cage structures are not unrelated.



	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>
Ginkgolide A	OH	H	H
Ginkgolide B	OH	OH	H
Ginkgolide C	OH	OH	OH
Ginkgolide M	H	OH	OH

Figure 19. Full structure of the four ginkgolides

We should first like to acknowledge our great indebtedness to the insight of the late Professor S. Fujise, who initiated investigations on these most fascinating compounds in our laboratory in 1960.

I am deeply indebted to my colleagues, Drs. M. Maruyama, A. Terahara, Y. Nakadaira, Y. Itagaki and Miss Y. Takagi for their very extensive studies, Dr. M. C. Woods (from Varian Associates) and I. Miura for the finer n.m.r. measurements, including the first application of NOE to structure determination, and to T. Dei, Y. Hirota, M. Miyashita and Y. Sugawara who participated in the studies during their undergraduate year. We are grateful to Dr. S. Sasaki for measurements of high-resolution mass spectra, and to Dr. Y. Inoue for valuable suggestions.

Thanks are due to the Takeda Pharmaceutical Industries for assistance in the large scale extraction of the ginkgolides, the Japan Electron Optics Laboratory for usage of the high-resolution mass spectrometer, and the Toyo Rayon Company for the grant to

## THE GINKGOLIDES

*purchase the Hitachi RMU-6D mass spectrometer. This investigation was supported by the National Institutes of Health, Public Health Service Research Grant No. CA08394.*

*Note added in proof.* Some aspects of the work reported in this article have since been published.<sup>16-20</sup>

## References

- <sup>1</sup> S. Furukawa. *Sci. Papers Inst. Phys. Chem. Res.* **26**, 178 (1935).
- <sup>2</sup> M. Sletzingel, and Ch. R. Dawson. *J. Org. Chem.* **14**, 670 (1946).
- <sup>3</sup> S. Furukawa. *Sci. Papers Inst. Phys. Chem. Res.* **24**, 314 (1934).
- <sup>4</sup> K. Hatano. *Nippon Rikagaku Kaishi* **38**, 425 (1950).
- <sup>5</sup> K. Ono. *Fukuoka Acta Med.* **47**, 731 (1956).
- <sup>6</sup> S. Furukawa. *Sci. Papers Inst. Phys. Chem. Res.* **19**, 39 (1932).
- <sup>7</sup> T. Kariyone, and H. Ageta. *Yakugaku Zasshi (Japan)* **19**, 58 (1959).
- <sup>8</sup> R. T. Major, P. Marchini, and T. Sproston. *J. Biol. Chem.* **235**, 3298 (1960). R. T. Major, P. Marchini, and A. J. Boulton. *J. Biol. Chem.* **238**, 1813 (1963). See also R. T. Major. *Science*, in press.
- <sup>9</sup> S. M. Manskaya, and L. A. Kodina. *Zhur. Prikladnoi Khimii* **32**, 2711 (1959).
- <sup>10</sup> K. Nakazawa. *Nippon Yakugaku Zasshi* **61**, 228 (1941).
- <sup>11</sup> W. Baker, A. C. M. Finch, W. D. Ollis, and K. W. Robinson. *Proc. Chem. Soc.* 91 (1959).
- <sup>12</sup> K. Nakazawa, and M. Ito. *Chem. Pharm. Bull. Japan* **11** (3), 283 (1963).
- <sup>13</sup> T. Kariyone, H. Kimura, and I. Nakamura. *Nippon Yakugaku Zasshi* **78**, 1152 (1958).
- <sup>14</sup> H. Kimura. *Nippon Yakugaku Zasshi* **82**, 214 (1962).
- <sup>15</sup> S. Winstein, P. Carter, F. A. L. Anet, and A. J. R. Bourn. *J. Am. Chem. Soc.* **87**, 5247 (1965);  
F. A. L. Anet, and A. J. R. Bourn. *J. Am. Chem. Soc.* **87**, 5250 (1965).
- <sup>16</sup> M. Maruyama, A. Terahara, Y. Itagaki, and K. Nakanishi. *Tetrahedron Letters* 299 (1967).
- <sup>17</sup> M. Maruyama, A. Terahara, Y. Itagaki, and K. Nakanishi. *Tetrahedron Letters* 303 (1967).
- <sup>18</sup> M. Maruyama, A. Terahara, Y. Nakadaira, M. C. Woods, and K. Nakanishi. *Tetrahedron Letters* 309 (1967).
- <sup>19</sup> M. Maruyama, A. Terahara, Y. Nakadaira, M. C. Woods, Y. Takagi, and K. Nakanishi. *Tetrahedron Letters* 315 (1967).
- <sup>20</sup> M. C. Woods, I. Miura, Y. Nakadaira, A. Terahara, M. Maruyama, and K. Nakanishi. *Tetrahedron Letters* 321 (1967).

## **SECTION 3: BIOSYNTHESIS**

### **INTRODUCTORY REMARKS BY THE HONORARY PRESIDENT OF THE SECTION**

**K. MOTHES**

It is an honour and a great pleasure for me to open this meeting of the section on Biosynthesis. The enormous development in the chemistry of natural products in the last twenty years has been combined with a major development in the biochemistry of these substances. Besides the new methods of preparing and analysing traces of compounds, the introduction of isotopes in feeding experiments helped to clear up relations between precursors and products of metabolism, and also clarified supposed reaction steps.

Probably the most important success of this new period of biochemistry may be seen in the fact that the fundamental metabolism of all known organisms is in principle the same: respiration, fermentation, protein synthesis, *etc.* This is the contribution of biochemistry to a general biology. The living world is a great unity.

There is, however, an enormous number of secondary products, which are not found in all organisms and which are, in the main, not necessary for life; they mostly have the character of excretes and are accumulated especially in plants, because this type of organism has no regularly working apparatus for clearing away the end products of metabolism. In the last ten years the biosynthesis of such secondary plant substances has been investigated with enormous success.

Feeding experiments with labelled compounds show what the plant is able to do but not always what it is doing within a normal metabolism. This means that those precursors with a good result are only possible precursors and not always natural ones. The plants have much more chemical potencies than they demonstrate under normal conditions. Queer relations may exist. There are, for instance, poppy species which change thebaine into codeine and morphine. Other papaver species are not able to do so. But there are fungi and tobacco species which easily reduce and demethylate thebaine to morphine.

The molecules are distinct forms of architecture. The investigations of biosynthesis will show us, without speculation, in which way nature constructs such molecules as distinct steric isomers. In his famous doctrine

#### K. MOTHES

about metamorphosis of plants Johann Wolfgang Goethe has endeavoured to visualize the manifold forms of higher plants as manifestations of a general type of a proto-plant. Were Goethe still alive he would find a more concrete field of comparative morphology. In the primary metabolism there seems to be realized only what is useful. But in the secondary plant substances nature realizes chemical ideas, changed a hundredfold, manifesting themselves in numerous different ways and yet remaining the same. The creation of secondary plant substances has, in general, nothing to do with usefulness. It rather is something to delight scientists, and sometimes to be used by man. For example, there are more than three hundred indole-alkaloids in the order of *Gentianales* which only consist of two building-stones: tryptophane and a terpene. This simple principle is repeatedly realized in an admirable manner.

Till now it was only possible to compare all these molecules formally, but now we begin to understand that this grand variety is ruled by a general law, by a congruency in biosynthesis. These two building stones can be found also in other combinations, for example in ergolines and emetines. One cannot but admire how few elements of construction Nature needs to develop the huge world of secondary products. Nature works as a great artist: simple in her means but inexhaustibly rich in imagination and application.

One of the scientists, who successfully elucidated these relations, not only with high chemical knowledge but also with a nature-inquiring imagination is Professor Battersby, who will give us a summary of some of his recent investigations. It is an honour and a great pleasure for me to introduce the first main speaker, Professor Battersby, on biosynthesis.

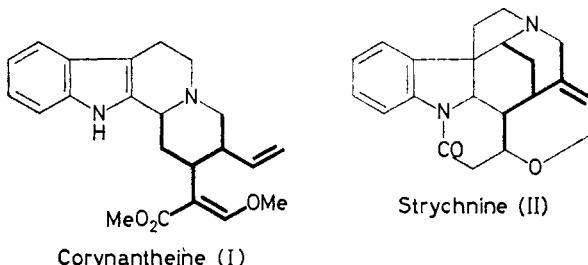
## BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

A. R. BATTERSBY

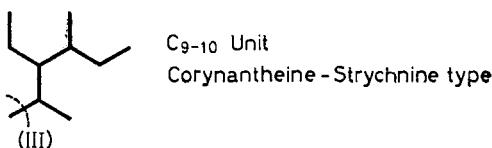
*The Robert Robinson Laboratories, University of Liverpool, Liverpool, U.K.*

There has been enormous progress over the past fifteen years or so in our understanding of the ways in which complex natural products are synthesized in living systems. In the field of alkaloids, as elsewhere, research during the initial phase<sup>1</sup> pin-pointed those substances which are used as the common building blocks, examples being acetic acid, ornithine and lysine for the reduced systems, and tyrosine, phenylalanine, 3,4-dihydroxyphenylpyruvic acid, and tryptophan for the many bases containing aromatic nuclei. Such knowledge allows the second phase of research to be undertaken; here the intermediates on the pathway are identified and this in turn leads to a study of the *mechanism* of each step in the biosynthetic sequence. At present a most exciting stage has been reached where hypothesis<sup>2-7</sup>, tracer experiment<sup>8,9</sup>, structure determination, and isolation work can be combined in a very powerful way. We can illustrate all these aspects by examining two quite different groups of alkaloids which have held our deep interest for several years. One is the large family of indole alkaloids, exemplified by corynantheine (I), strychnine (II), catharanthine (IV, Figure 1), and vindoline (VII), and the other contains the alkaloids found in *Colchicum* species of which the ancient poison colchicine (XXXVI) is the best known member. The origins of these groups of alkaloids have remained until very recently the two major unsolved problems in the field.

The number of known indole alkaloids has increased sharply over the last few years and there are now about six hundred<sup>10,11</sup>. A tryptamine



Corynantheine (I)



C<sub>9-10</sub> Unit  
Corynantheine - Strychnine type

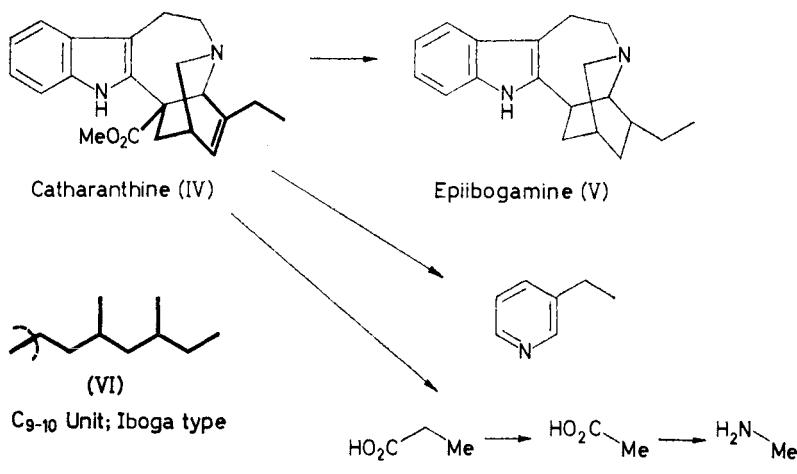


Figure 1

residue (normal bonds in I, IV, and VII) appears almost invariably and in the few cases examined by tracer methods<sup>12</sup>, this residue has been found to be derived in the expected way from tryptophan. The remaining nine or ten skeletal carbon atoms (thickened bonds in I, IV, and VII) appear in what at first sight seems a bewildering variety of different arrangements but closer inspection allows three main groups to be discerned<sup>13</sup>. Together, these three main groups account for the vast majority of indole alkaloids. We can conveniently refer to them as: (a) the corynantheine-strychnine

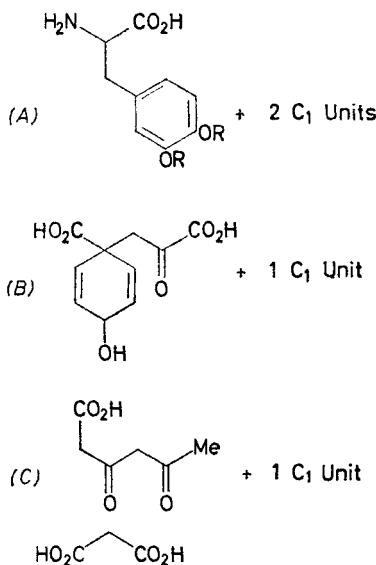
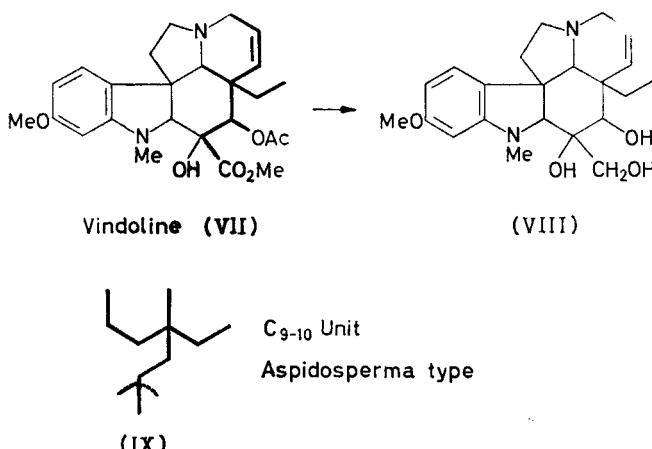


Figure 2

BIOSYNTHESIS OF THE INDOLE AND COLCHICUM ALKALOIDS



type which possess the C<sub>9-10</sub> unit (III), (b) the *Aspidosperma* type having the C<sub>9-10</sub> unit (IX), and (c) the *Iboga* type where the C<sub>9-10</sub> unit appears as (VI). In those alkaloids where only nine skeletal carbons appear in addition to the tryptamine residue, it is invariably the carbon atom indicated by the dotted line which has been lost.

Our own experimental study of the origin of this ubiquitous C<sub>9-10</sub> unit started some six years ago and all the early work was designed to test rigorously three hypotheses<sup>7,14,15</sup>; these are illustrated in *Figure 2* simply by the units involved. Our results have been published<sup>16-18</sup> together with complementary work<sup>19-21</sup> and it is only necessary therefore to summarize the main findings. It was proved beyond doubt that all three hypotheses are incorrect; the earlier experimental support<sup>15,22</sup> for the hypothesis (C) [in *Figure 2*] has since been withdrawn<sup>23</sup>. Further, the results established that a C<sub>1</sub>-unit is *not* involved in the biosynthesis. The only hypothesis in accordance with this finding is that due to Thomas<sup>24</sup> and Wenkert<sup>7</sup>, which suggested a relationship of the C<sub>9-10</sub> unit to the cyclopentane monoterpene skeleton (XII). *Figure 3* shows how the corynantheine-strychnine C<sub>9-10</sub> unit (III) could be derived by cleavage of the cyclopentane ring of some unknown monoterpene illustrated here simply as a skeleton without any indication of its oxidation level. One can extend the structural relationships by observing that the C<sub>9-10</sub> units of the *Aspidosperma* and *Iboga* types may be derived by a combination of a further bond fission as indicated and a bond formation in the direction either (a) or (b). It must be emphasized that *Figure 3* is not intended to convey any information about the *timing* of the various changes. This scheme simply means that the three types of C<sub>9-10</sub> units may be so related by steps occurring at points, as yet unknown, somewhere along the biosynthetic pathway.

A test of the monoterpene theory clearly involves feeding experiments with sodium mevalonate (X) and our first small incorporations of this precursor were achieved into the alkaloids of *Cephaelis ipecacuanha*† and

†This plant contains the isoquinoline relatives of the indole alkaloids

## A. R. BATTERSBY

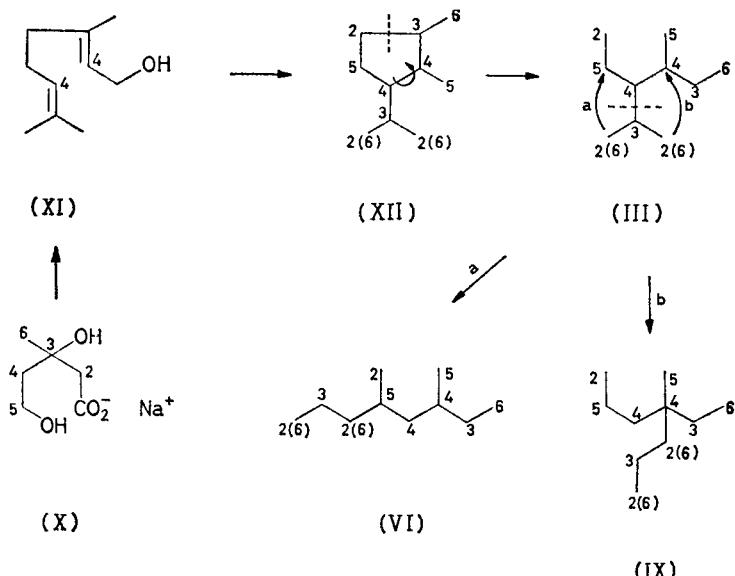


Figure 3

*Rauvolfia serpentina*<sup>18</sup>. Other plants, particularly *Rhazia stricta* and *Vinca rosea* were better able to draw external supplies of sodium mevalonate into their biosynthetic systems and gave satisfactory incorporations of activity<sup>25</sup>. Thus, sodium [ $2\text{-}^{14}\text{C}$ ]mevalonate fed to *R. stricta* yielded radioactive 1,2-dehydroaspidospermidine (XVII; 0.15 per cent incorporation) and similarly, active vindoline (VII; 0.5 per cent incorporation) was isolated from *V. rosea* plants. The latter plant also afforded radioactive serpentine (XIII), ajmalicine (XV), catharanthine (IV) and perivine (XIX). Simultaneously and quite independently, Professors A. I. Scott and D. Arigoni and their respective coworkers carried out parallel experiments<sup>26,27</sup>. All the results which follow were obtained in our Robert Robinson Laboratory and our colleagues' work will be drawn in at the appropriate points by indicating on the various Figures which results they also have obtained; this will be done by A<sup>27</sup> and S<sup>26</sup> (for Arigoni and Scott respectively), the superscripts being literature references.

In our own experiments, it was decided to determine the labelling pattern generated in each of the three types of alkaloid by feeding the plants with various sodium mevalonates carrying specific <sup>14</sup>C-labels. [ $2\text{-}^{14}\text{C}$ ]Mevalonolactone is available commercially and specimens of [ $3\text{-}^{14}\text{C}$ ]-, [ $4\text{-}^{14}\text{C}$ ]-, and [ $5\text{-}^{14}\text{C}$ ]mevalonolactone were synthesized<sup>28,29</sup>; these were utilized in separate feeding experiments with *Rhazia stricta* and *Vinca rosea* plants. A further essential requirement was the development of unambiguous methods for degrading the labelled alkaloids and the procedures used are shown in Figures 1 and 4-7. In the design of the various sequences, we drew upon previous work, particularly the base cleavage of serpentine (XIII) to afford harman<sup>12</sup> (XIV) (Figure 4) the hydrolysis and decarboxylation of

BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

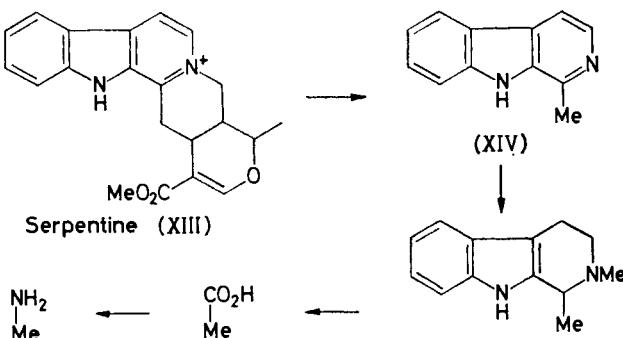


Figure 4

ajmalicine (XV) followed by Wolff-Kishner reduction of the product to yield ajmaliciol<sup>7</sup> (XVI) (Figure 5), and the conversion of catharanthine (IV) by hydrogenation and hydrolysis into epiibogamine<sup>30</sup> (V) (Figure 1). The subsequent steps, especially the Kuhn-Roth and Schmidt degradations (which were also used to examine the ethyl side-chain of vindoline) were rigorously controlled by the use of standard substances at a known level of radioactivity. Finally, it was necessary to devise a suitable method for the isolation of the carbon atom marked 2 in 1,2-dehydroaspidospermidine (XVII) (Figure 6). Reduction of the alkaloid with lithium aluminium hydride<sup>31</sup> and protection of the indolinic  $>\text{NH}$  by acetylation allowed Emde and Hofmann degradations to be carried out as for aspidospermine<sup>32</sup>. Cleavage of the olefin with osmium tetroxide-periodate then afforded the required carbon as formaldehyde. The degradation of perivine is shown in Figure 7.

These then are the methods used to degrade the various alkaloids isolated from each feeding experiment and the results are collected in Figure 8. This illustrates the patterns expected on the basis of head-to-tail combination of two  $\text{C}_5$  units (derived from mevalonate) to yield a cyclopentane system

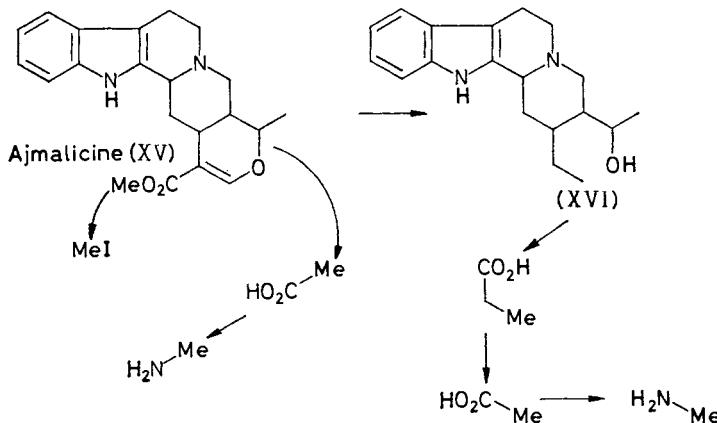


Figure 5

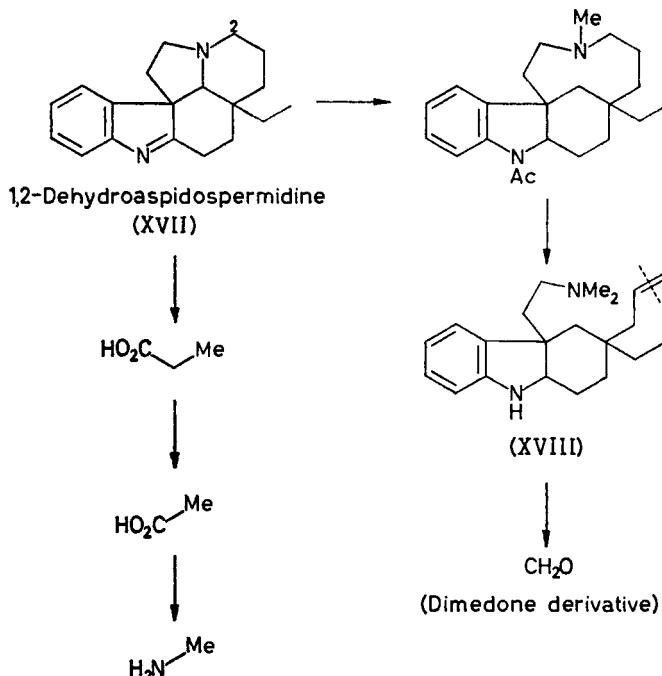


Figure 6

and subsequent cleavage of it. The underlined numbers (in Figure 8) indicate the positions which have been proved to be so labelled from the appropriately labelled sodium mevalonate. The quantitative results agreed closely with the theoretical values. These results lead to several important conclusions and we shall concentrate initially on the corynantheine-strychnine type of C<sub>9-10</sub> unit. The results prove that a bond has been formed during the biosynthesis between the C-4 carbons of the two mevalonate residues in agreement with cyclopentane ring-closure. Further, a labelling pattern which is almost complete has been derived for the corynantheine-strychnine unit and the pattern is quantitatively in agreement with the illustrated scheme; one should notice that C-2 and C-6 of one mevalonate

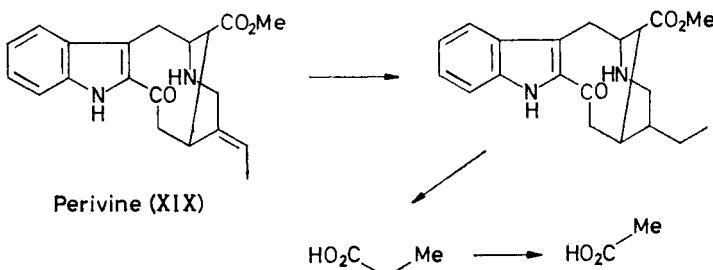


Figure 7

BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

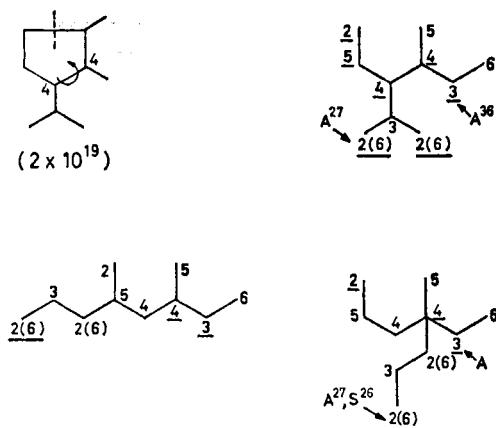


Figure 8

unit became equivalent during the biosynthesis as was found earlier for the cyclopentane monoterpene plumeride<sup>33</sup>. Finally, the labelling patterns derived so far for the *Iboga* and *Aspidosperma* types of C<sub>9-10</sub> unit are in complete agreement with their suggested derivation at some stage from the corynantheine-strychnine unit in ways we have already considered.

These results taken as a whole constitute powerful evidence supporting the formation of a cyclopentane monoterpene skeleton during the biosynthesis of indole alkaloids but it was important to obtain further evidence that a head-to-tail C<sub>10</sub> unit is involved. Bearing in mind present knowledge of the biochemical transformation of mevalonic acid, one would expect geraniol (XI), or a derivative thereof, to act as a precursor of the cyclopentane system. Accordingly, [2-<sup>14</sup>C]geraniol was prepared by known methods<sup>34</sup> and was converted largely into the corresponding pyrophosphate. This was incorporated into all three types of indole alkaloids in *Vinca rosea* plants, the yield being four to thirty times higher than had been obtained

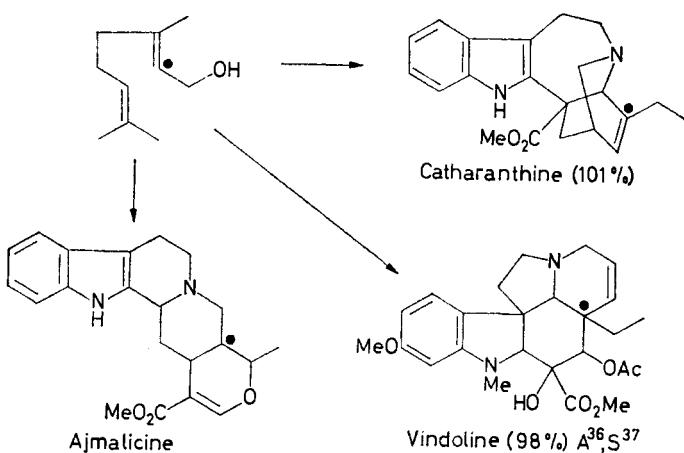
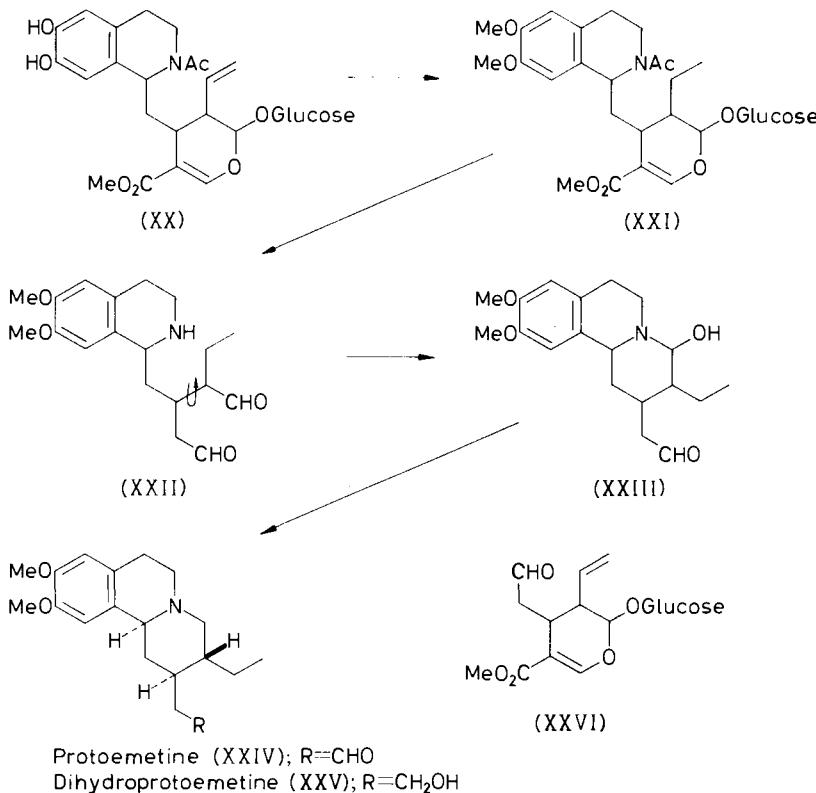


Figure 9

A. R. BATTERSBY

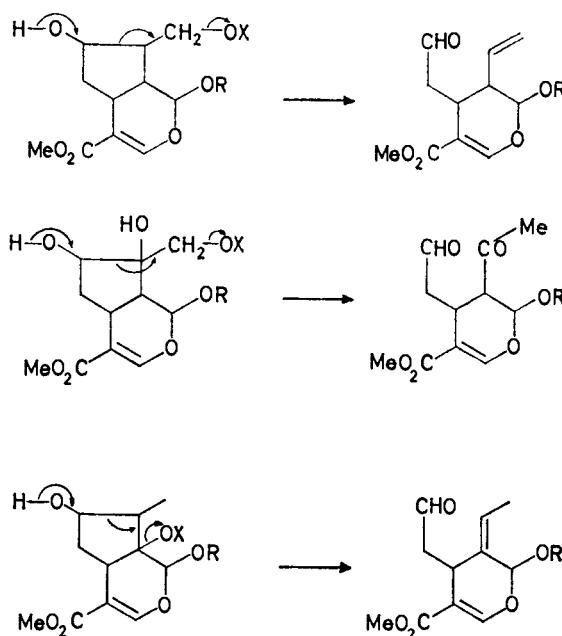
from sodium mevalonate. The label of geraniol had been selected to allow ready isolation of the radioactive carbons from the various alkaloids (see *Figure 9*). Thus, for vindoline (as its desacetyl derivative) and catharanthine, Kuhn-Roth degradation afforded propionic acid of the same molar activity as the original alkaloid (see percentages in *Figure 9*) together with acetic acid which was radio-inactive. All the activity of these alkaloids is therefore located at the indicated positions<sup>35</sup>, both results being entirely consistent with the cyclopentane monoterpene scheme (*Figure 3*). Work is still in progress on ajmalicine and the present position is that six of the ten carbon atoms of the C<sub>10</sub> unit have been shown to carry no activity, so limiting the label to C-3, 14, 20 and 21; the expected site of labelling is C-20.

We can now attempt with some confidence to discover the nature of the cyclopentane monoterpenes (for it is probable that there are several) which lie on the biosynthetic pathway to the indole alkaloids. If one considers the various possible oxidation levels of the ten carbon atoms of a cyclopentane monoterpene skeleton it turns out that there are  $2 \times 10^{19}$  possible structures. We were greatly helped in reducing this number by the structural work on ipecoside which we have been carrying out in collaboration with Professors Janot and Levisalles. This substance is a glucoside occurring in *Ipecacuanha* plants which also biosynthesize the isoquinoline analogues of the indole alkaloids. Thus, the *Ipecacuanha* alkaloid



## BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

protoemetine<sup>38</sup> (XXIV) should be compared with corynantheine (I). It is not possible here to cover the extensive chemical and spectroscopic work which led us to consider the constitution (XX) for ipecoside; it will suffice to say that this structure accommodates all the experimental findings. To establish this constitution beyond doubt and at the same time to determine the complete absolute stereochemistry, we decided to correlate ipecoside with protoemetine (XXIV). Acidic hydrolysis cleaved the *N*-acetyl group and the ester function to allow decarboxylation of the  $\beta$ -aldehydo acid. The mixture of bases so obtained was expected to contain the carbinolamine (XXIII) which should be susceptible to reduction by dissolving metals.

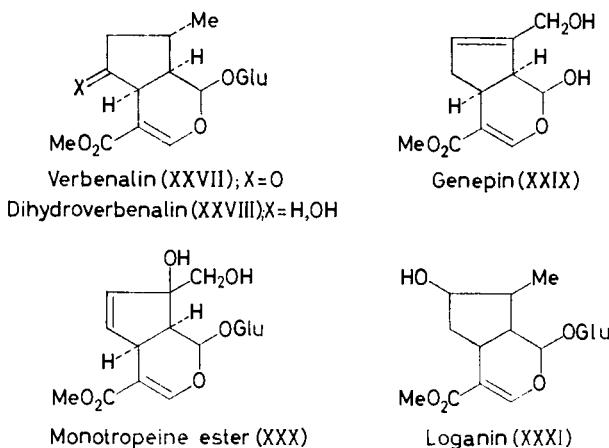


*Figure 10*

Zinc-sulphuric acid reduction afforded a mixture from which dihydroprotoemetine (XXV) was isolated. The constitution (XX) is therefore established in every detail and ipecoside stands as the first example of a mixed isoquinoline-monoterpene alkaloid.

We can now speculate that the related base (XX,  $>\text{NH}$  in place of  $>\text{NAC}$ ) is the biosynthetic precursor of protoemetine and that ipecoside arises by a nitrogen-blocking acetylation. If this is true, the substance which combines with dopamine to afford desacetylipecoside could reasonably be the aldehyde (XXVI) or its equivalent; we are now close to a cyclopentane monoterpene. How might such an aldehyde (XXVI) be generated by suitable fragmentation of the cyclopentane ring? *Figure 10* sets out some reasonable possibilities where X might be a phosphate residue to provide a good leaving group. These hypothetical precursors can be

compared with the four natural monoterpenes which they most nearly resemble, verbenalin<sup>39</sup> (XXVII), genepin<sup>40</sup> (XXIX), monotropeine<sup>41</sup> shown as the methyl ester (XXX), and loganin<sup>42</sup> (XXXI). The last was *a priori* the most attractive in that it is structurally closest to the hypothetical precursors and, moreover, occurs alongside indole alkaloids in *Strychnos* species<sup>42</sup>. All four compounds (XXVII), (XXIX), (XXX) and (XXXI)



were tested in plant feeding experiments with *Vinca rosea* and the work on all but (XXIX) is complete. Verbenalin (XXVII) and dihydroverbenalin (XXVIII) were labelled by tritium exchange and were found to be totally ineffective as precursors of the indole alkaloids. Similarly, monotropeine methyl ester (XXX), tritium labelled at the ester methyl gave zero incorporation into the alkaloids. This result not only points strongly against the monotropeine system as the correct monoterpene but also shows, most

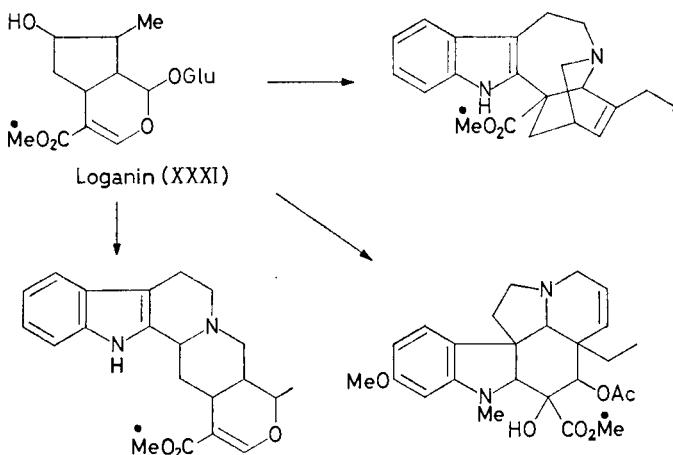
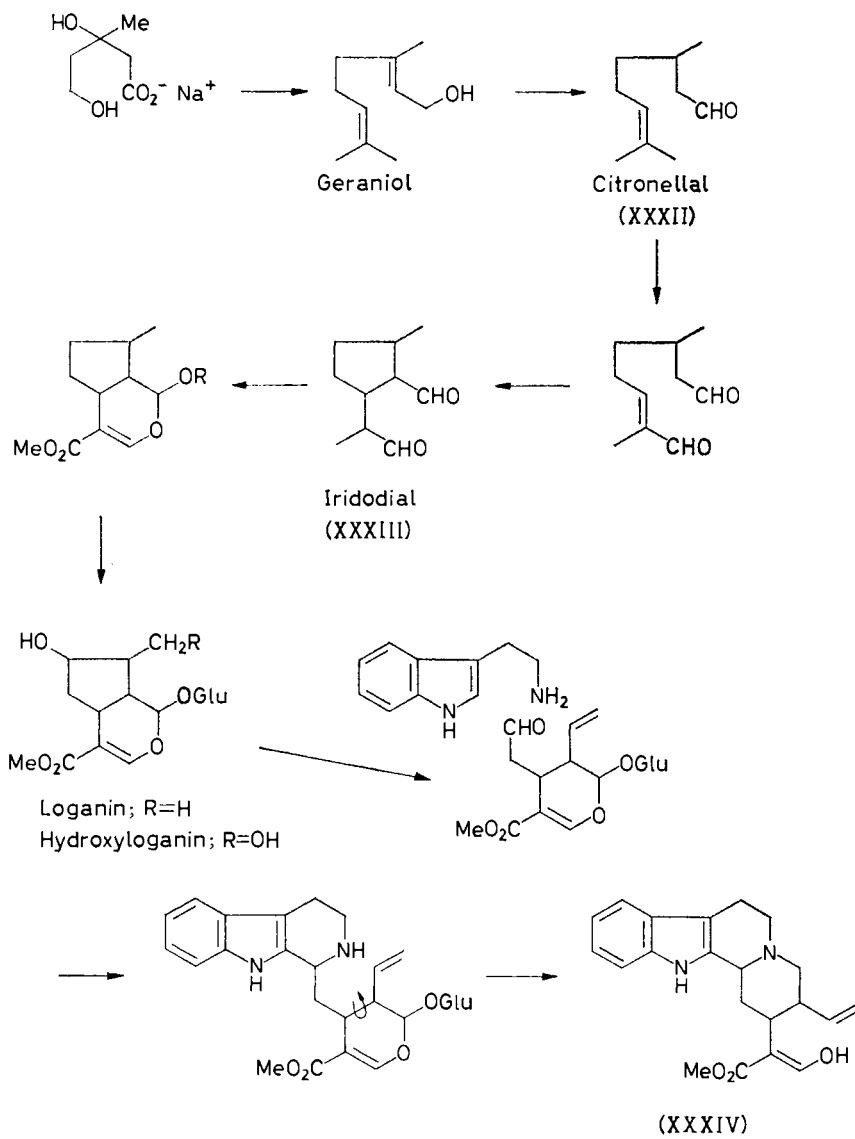


Figure 11

BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

importantly, that the labelled methyl group does not undergo transfer into the methylating system of the plant. Thus, when loganin (XXXI), also  $^3\text{H}$ -labelled at the methyl ester group afforded good incorporations of activity into all three types of indole alkaloid [*ca.* 1 per cent into vindoline (VII)] we felt confident that the explanation was not a trivial one of methyl transfer (*Figure 11*). This was put beyond doubt by converting the active vindoline first into desacetyl vindoline (VII, OH in place of OAc), without loss of activity, and then reducing desacetylvinodine to the diol (VIII). Less than 0·1 per cent of the original activity was retained in the diol proving



*Figure 12*

that the ester methyl group carries all the activity and that none is present in the  $>\text{NMe}$  or aryl OMe groups; this result eliminates methyl transfer. Both ajmalicine and catharanthine (see Figure 11) were also highly radioactive and were proved to be labelled solely at their *O*-methyl groups. So the supreme importance of the cyclopentane monoterpene system for the biosynthesis of the indole alkaloids is established and loganin (XXXI) is identified as one of the key intermediates.

The pathway to the large and varied family of indole alkaloids can now take real form and Figure 12 shows an attractive sequence†. Some parts of the scheme rest on the firm experimental basis described so far in this lecture and the rest are under intensive study. Very recent experiments indicate that the conversion of the corynantheine–strychnine  $\text{C}_{9-10}$  unit into the *Aspidosperma* and *Iboga*  $\text{C}_{9-10}$  units occurs after introduction of the nitrogen but further work is necessary here; this is in progress.

The problem is at a most fascinating stage where the researcher can see that the precise detail of the pathway to the indole alkaloids cannot now escape him.

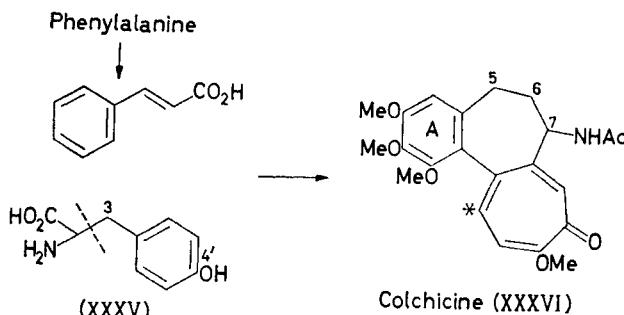


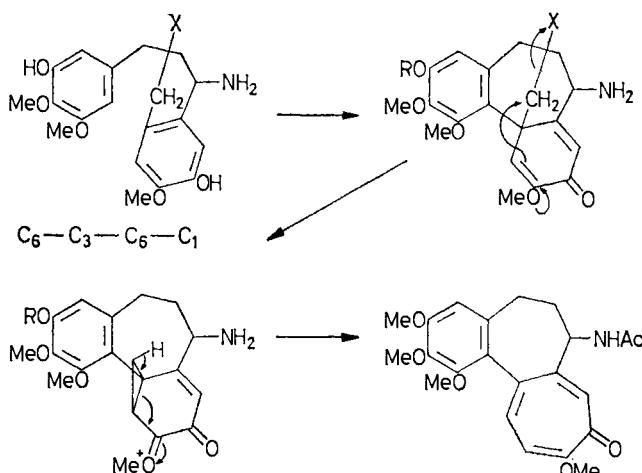
Figure 13

Let us now turn to the other problem, that posed by colchicine (XXXVI) and its relatives. One must determine the origin of the tropolone ring and also explain how the nitrogen atom comes to be placed in this unusual position; the solution here has proved to be a great surprise. The early tracer work established that ring A of colchicine and the atoms 5, 6 and 7 are derived from phenylalanine by way of cinnamic acid by steps which were not understood at that stage<sup>43-45</sup>. We further found that when [ $3-\text{14C}$ ] tyrosine was fed to *Colchicum autumnale* plants, it was incorporated well into colchicine to label the tropolone ring specifically at the starred position (Figure 13) and on this basis we put forward<sup>45,46</sup> the biosynthetic scheme shown in Figure 14. Here the suggestion is that a  $\text{C}_6-\text{C}_3$  residue from cinnamic acid is combined with a  $\text{C}_6-\text{C}_1$  unit arising from tyrosine, to afford a  $\text{C}_6-\text{C}_3-\text{C}_6-\text{C}_1$  system. Phenol oxidation<sup>5,6</sup> could then generate the dienone

†We must emphasize that several closely similar schemes could be written in which the sequence of operations is altered. For example, though at present it is attractive to consider cleavage of the cyclopentane ring before the nitrogenous portion of the molecule is introduced, the evidence is indirect. Plausible schemes reversing the order can be written. Appropriate experiments are in hand.

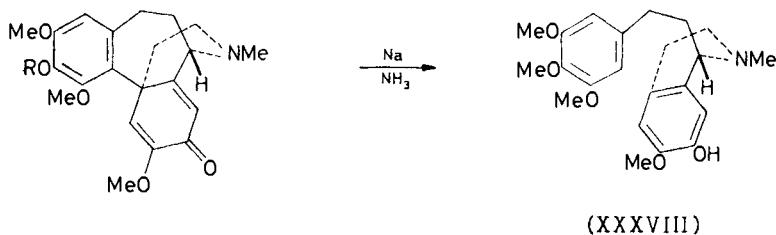
## BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

in which X is some good leaving group, possibly, we thought, a phosphate ester. Homoallylic assistance<sup>47</sup> of the separation of X could then bring about ring expansion as illustrated<sup>46</sup>. The label from [3-<sup>14</sup>C]tyrosine would appear by this process at the correct position in the tropolone ring. This idea was supported by Leete<sup>48</sup> who fed [4'-<sup>14</sup>C]tyrosine (see XXXV) to the plants; the isolated colchicine was labelled only at the carbonyl carbon of the tropolone ring. The C<sub>6</sub>-C<sub>3</sub>-C<sub>6</sub>-C<sub>1</sub> "precursor" with X as OH (Figure 14) was in fact synthesized in labelled form at Liverpool and was fed to autumn crocus plants but there was no incorporation of activity; this is as it should be as will become clear in the sequel.

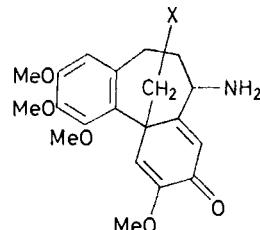


*Figure 14*

Here again the tracer studies were able to leap forward as a result of crucial information obtained in related structural studies. *Androcymbium melanthioides* is closely related to the autumn crocus and it contains several alkaloids<sup>49</sup>, among them colchicine and one of unknown structure named androcymbine. As for ipécacuine, it is outside the scope of this lecture to describe how the structure and absolute stereochemistry (XXXVII, R = H) were determined<sup>50</sup> (in joint work with Professor F. Šantavý). The key reaction was the reductive cleavage<sup>51</sup> of *O*-methylandrocymbine (XXXVII, R = Me) to afford the 1-phenethylisoquinoline (XXXVIII) which was synthesized<sup>50</sup>. Not only was androcymbine of great interest by its being the first 1-phenethylisoquinoline alkaloid to be discovered, but also because of its relationship to dienone (XXXIX), the hypothetical "precursor" of colchicine. Moreover, the absolute stereochemistry of androcymbine is the same as that of colchicine. It was clear that all the available information would fall perfectly into place if colchicine is biosynthesized from a 1-phenethylisoquinoline precursor. On this basis, structure (XXXIX) derived directly from tracer experiments requires extension to that of *O*-methylandrocymbine (XXXVII, R = Me). The biosynthetic scheme shown in *Figure 15* can now be considered for colchicine and for the base



Androcymbine (XXXVII); R=H

[ *A. melanthioides* ]

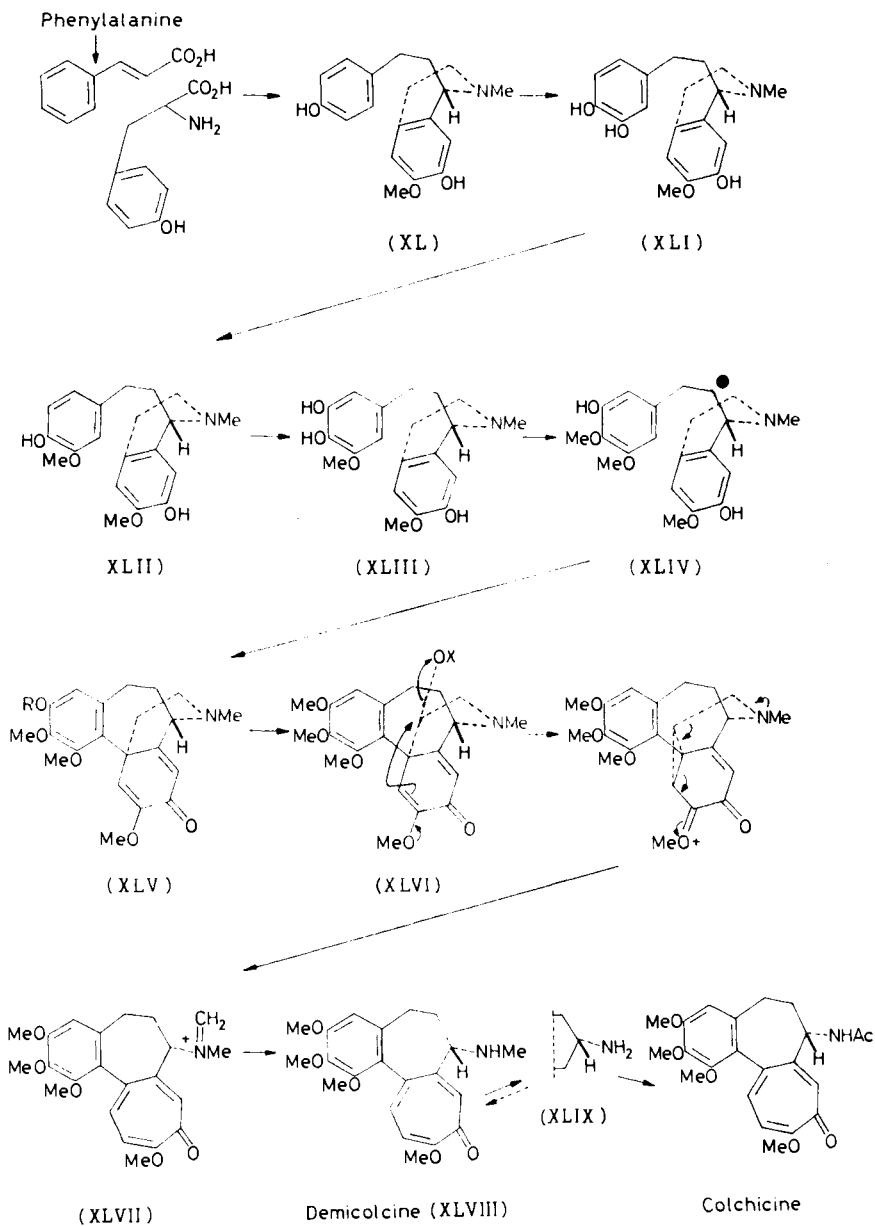
(XXXIX)

which also occurs in *Colchicum* species, demecolcine (XLVIII). The first few steps are best considered later and the present consideration can start at the diphenolic 1-phenethylisoquinoline (XLIV). Phenol oxidation, methylation of the dienone to yield *O*-methylandrocymbine (XLV; R = Me), and hydroxylation to form (XLVI; X = H) need no comment and the homoallylic assistance of ionization has been considered earlier. In this case, electron release from nitrogen can satisfy the deficiency on oxygen and models indicate that this process should be a favourable one. The resultant imine (XLVII) could then readily hydrolyse to afford demecolcine (XLVIII) which on this view precedes colchicine in the biosynthetic process.

Several crucial tests of this scheme can be made. Firstly, it collapses if *O*-methylandrocymbine (XLV; R = Me) is not a precursor of colchicine. In the experimental test, it was gratifying to find that this substance is incorporated without randomisation of the label into colchicine (*Colchicum autumnale*) with the high yield of over 15 per cent. Further, the isoquinoline (XLIV) was synthesized <sup>14</sup>C-labelled at the indicated position and was used by the plants to make colchicine in over 10 per cent yield; degradation showed the colchicine to be labelled solely at the expected site. These two experiments laid open the problem of colchicine biosynthesis and we can now turn to researches which probed the detailed mechanism of this remarkable pathway. Much of this can be carried out by using the correctly labelled forms of the important 1-phenethylisoquinoline (XLIV), since a carbon-14 label at position 6 is now known to be retained and thus can act as an internal standard for the experiments which follow.

If the pathway outlined in Figure 15 is correct, then several strict requirements are imposed upon the incorporation of (XLIV) into colchicine and the twelve labels shown in Figure 16 are required to test these points. Of course, use of a dodecatuply labelled substance in plant feeding experiments

BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS



*Figure 15*

would make the problem of degrading the isolated colchicine an extremely difficult one and therefore the twelve labels were grouped into convenient sets for separate administration to the plants. Several different syntheses of this precursor were required to allow labelling to be achieved at the various required points but it is not necessary here to cover this aspect.

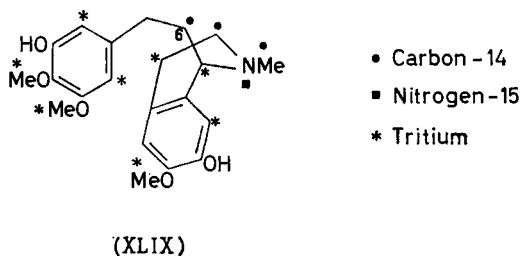


Figure 16

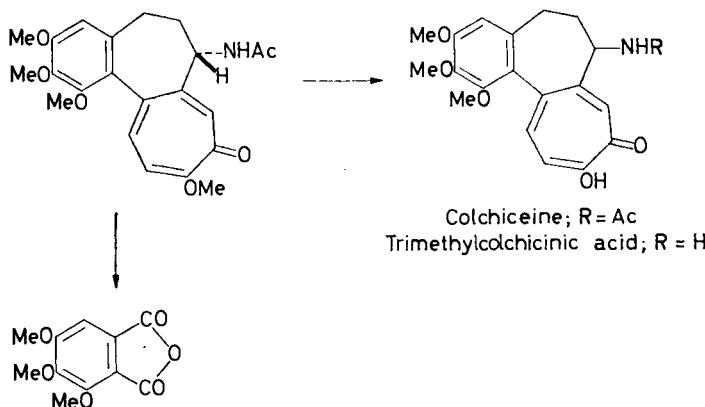
The various requirements of the proposed biosynthetic pathway (*Figure 15*) can now be considered in turn together with the results gained so far from the multiply labelled precursor (XLIX). (a) The nitrogen atom must be retained. This is in sharp contrast to most of the earlier ideas on colchicine biosynthesis in which the nitrogen was thought to be introduced after the main skeleton had been constructed. The  $^{15}\text{N}:\text{C}^{14}$  experiment showed that the dilution of  $\text{C}^{14}$  from precursor (XLIX) to colchicine was 157 and this matched perfectly the dilution of  $^{15}\text{N}$  which was 156. The nitrogen atom of colchicine is thus proved to be that of the original isoquinoline. The  $^{15}\text{N}$  analysis was carried out at Saskatoon by Drs. E. W. Underhill and L. R. Wetter and we are most grateful to these Canadian colleagues. (b) One of the three  $^3\text{H}$  atoms attached directly to the aromatic nuclei of (XLIX) should be lost and the  $^3\text{H}:\text{C}^{14}$  ratio found in the isolated colchicine proved this to be precisely correct. (c) All three methoxyl groups should remain intact and here some simple degradations (*Figure 17*) of the colchicine were required to determine the necessary  $^3\text{H}:\text{C}^{14}$  ratios. These established that no significant loss of any of the methoxyl groups occurs during the biosynthesis.

We can now turn to the mechanism of the ring expansion step. One would expect that direct enzymatic hydroxylation of *O*-methylandrocymbine (XLV, R = Me) to form the hydroxy derivative (XLVI, X = H), would be a stereospecific one and this should remove half the tritium present at this position in the labelled precursor (XLIX). Obviously if this carbon reaches the carbonyl state of oxidation, all the tritium will be lost whereas non-stereospecific oxidation will lead to a retention of some 85 per cent of the original tritium due to the  $^3\text{H}$  isotope effect; this we know from other tracer experiments<sup>52</sup>. The loss of tritium found for colchicine biosynthesis was 56 per cent in clear agreement with a stereospecific hydroxylation reaction. Whether the difference from 50 per cent is significant is not yet known since this is near the limit of accuracy for the determination of  $^3\text{H}:\text{C}^{14}$  ratios. The main point, however, is proved and an answer to this secondary one must wait the outcome of more refined experiments now in hand.

The mechanism requires that the bridge carbon atom adjacent to the nitrogen of (XLIX) must be cleaved off in the formation of colchicine and here the ratios in the multiply labelled experiment showed that *ca.* 85 per cent had indeed been lost. This difference from the expected value (100 per cent) is greater than the experimental error of counting and it was necessary

## BIOSYNTHESIS OF THE INDOLE AND *COLCHICUM* ALKALOIDS

to show by conversion of the colchicine into trimethylcolchicinic acid (*Figure 17*) that no significant activity was present in the *N*-acetyl group. The results so far establish that the bridge carbon is completely lost in accord with the scheme but a little of the  $^{14}\text{C}$  activity it carries is probably trapped in the plant's metabolic pools (particularly the one carbon and two carbon pools). A small part of the avitivity could be passed back in this way rather generally into the colchicine then being synthesized. This would account well for our findings.



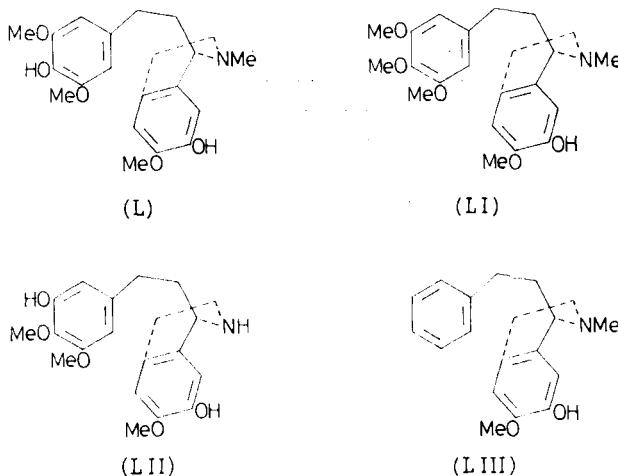
*Figure 17*

A similar explanation probably also holds for the values obtained when the plants were fed with (XLIX) labelled at the *N*-methyl group and carrying appropriate labels as internal standards. Here, the *N*-methyl group was fully retained during the biological conversion of (XLIX) into demecolcine (XLVIII) in accord with the scheme but for colchicine, the fall in  $^{14}\text{C}$  activity corresponded to an elimination of *ca.* 90 per cent of the *N*-methyl label. Since the *N*-acetyl group of colchicine was shown not to be significantly labelled in this experiment, a small general feed-back of activity is the most probable explanation. Of course, the extent of this feed-back can be determined experimentally, and this is in hand.

A study of the final steps of the biosynthetic scheme (*Figure 15*) provided valuable further evidence. [*O*-methyl- $^3\text{H}$ ]Demecolcine (XLVIII) was converted into colchicine to the high extent of 14 per cent in *Colchicum autumnale* plants and there was only minor conversion of labelled colchicine into demecolcine, in this case using *C. byzantinum*. Further, a very satisfactory chemical method was devised for the preparation of desacetyl-colchicine (XLIX) from colchicine which allowed the preparation of the labelled desacetyl base (XLIX). When this was administered to the plants, it was incorporated very well into colchicine (38 per cent) and also to a much smaller extent (3.6 per cent) into demecolcine. Clearly the forward path demecolcine  $\rightarrow$  desacetylcolchicine  $\rightarrow$  colchicine is the major one.

The researches which have been outlined define in considerable detail the pathway from the diphenolic base (XLIV) to demecolcine and colchicine. We can go further than this. The various 1-phenethylisoquinolines (XL),

A. R. BATTERSBY



(XLII), and (XLIII) have all been synthesized carrying skeletal  $^{14}\text{C}$ -labels and they are all incorporated into colchicine to extents entirely consistent with the illustrated sequential build up of oxygenation†.

The pathway for colchicine and its relatives is thus known almost completely and it is a very surprising one. Colchicine now falls satisfactorily into place, rather than being an "odd man out" in the alkaloid field. Every organic chemist will admire the beauty of the natural pathway and a synthesis of colchicine by this sequence is a tempting possibility. The necessary experimental work is in progress.

*The work I have described herein could only have been achieved by an enthusiastic and vigorous team effort and I cannot praise my colleagues too highly. Drs. R. T. Brown, B. Gregory, R. S. Kapil, J. A. Knight, J. A. Martin, and A. O. Plunkett carried out all the recent work on the indole and Ipecacuanha alkaloids. The group researching on colchicine and its relatives comprised A. Barker, J. H. Clements, E. McDonald, and Drs. R. B. Herbert and R. Ramage. In addition to his own work in the colchicine area, Dr. Ramage and my other senior colleague, Dr. J. Staunton, made innumerable contributions to the general good of the work.*

*Many friends helped us with rare alkaloids and terpenes and we are greatly indebted for such gifts to Drs. N. Neuss, H. T. Openshaw, G. F. Smith, and W. I. Taylor and to Professors G. Büchi, G. W. K. Cavill, C. Djerassi, H. Inouye, E. Ramstad, E. R. Ritchie, and F. Santavy. We record our warmest thanks to these colleagues.*

†The bases (L), (LI), (LII), and (LIII) were also prepared in labelled form. In one way or another, these are the wrong "precursors" bearing in mind the reasoning outlined in this lecture. For example, (LI) and (LIII) lack phenolic groups in the correct positions to permit the phenol coupling step. All these substances failed to act as precursors of colchicine in the autumn crocus.

## BIOSYNTHESIS OF THE INDOLE AND COLCHICUM ALKALOIDS

### References

- <sup>1</sup> Reviewed *inter al* by A. R. Battersby. *Quart. Rev.* **15**, 259 (1961); K. Mothes and H. R. Schütte. *Angew. Chem.* **75**, 265, 357, (1963).
- <sup>2</sup> R. Robinson. *The Structural Relations of Natural Products*, Clarendon Press, Oxford 1955, and earlier references cited therein.
- <sup>3</sup> C. Schöpf. *Naturwiss.* **39**, 241 (1952), and earlier references cited therein.
- <sup>4</sup> R. B. Woodward. *Angew. Chem.* **68**, 13 (1956).
- <sup>5</sup> D. H. R. Barton and T. Cohen. *Festschrift A. Stoll*, Birkhauser, Basle, 1957, p.117.
- <sup>6</sup> H. Erdtman and C. A. Wachtmeister. *Festschrift A. Stoll*, Birkhauser, Basle, 1957, p. 144.
- <sup>7</sup> E. Wenkert and N. V. Bringi. *J. Am. Chem. Soc.* **81**, 1474 (1959); E. Wenkert. *J. Am. Chem. Soc.* **84**, 98 (1962).
- <sup>8</sup> A. R. Battersby. Tilden Lecture, *Proc. Chem. Soc.* 189 (1963).
- <sup>9</sup> D. H. R. Barton. Hugo Müller Lecture. *Proc. Chem. Soc.*, 293 (1963).
- <sup>10</sup> *The Alkaloids*, Vol. VIII, Ed. R. H. F. Manske, Academic Press, New York (1965).
- <sup>11</sup> M. Hesse. *Indolalkaloide in Tabellen* Springer-Verlag, Berlin (1964).
- <sup>12</sup> *E.g.*, E. Leete. *Tetrahedron* **14**, 35 (1961).
- <sup>13</sup> E. Schlitter and W. I. Taylor. *Experientia* **16**, 244 (1960).
- <sup>14</sup> R. B. Woodward. *Nature* **162**, 155 (1948).
- <sup>15</sup> E. Leete, S. Ghosal, and P. W. Edwards. *J. Am. Chem. Soc.* **84**, 1068 (1962).
- <sup>16</sup> A. R. Battersby. *Biogenesi delle Sostanze Naturali*, Accademia Nazionale dei Lincei, Roma, 1964, p. 57.
- <sup>17</sup> A. R. Battersby, R. Binks, W. Lawrie, G. V. Parry, and B. R. Webster. *Proc. Chem. Soc.* 369 (1963).
- <sup>18</sup> A. R. Battersby, R. Binks, W. Lawrie, G. V. Parry, and B. R. Webster. *J. Chem. Soc.* 7459 (1965).
- <sup>19</sup> H. Goeggel and D. Arigoni. *Experientia* **21**, 369 (1965).
- <sup>20</sup> D. H. R. Barton, G. W. Kirby, R. H. Prager, and E. M. Wilson. *J. Chem. Soc.* 3990 (1965).
- <sup>21</sup> K. Stolle, D. Gröger, and K. Mothes. *Chem. and Ind.* 2065 (1965).
- <sup>22</sup> P. N. Edwards and E. Leete. *Chem. and Ind.* 1666 (1961); E. Leete and S. Ghosal. *Tetrahedron Letters*, 1179 (1962).
- <sup>23</sup> E. Leete, A. Ahmad, and I. Kompis, *J. Am. Chem. Soc.* **87**, 4168 (1965).
- <sup>24</sup> R. Thomas. *Tetrahedron Letters* 544 (1961).
- <sup>25</sup> A. R. Battersby, R. T. Brown, R. S. Kapil, A. O. Plunkett, and J. B. Taylor. *Chem. Comm.* 46 (1966).
- <sup>26</sup> F. McCapra, T. Money, A. I. Scott, and I. G. Wright. *Chem. Comm.* 537 (1965).
- <sup>27</sup> H. Goeggel and D. Arigoni. *Chem. Comm.* 538 (1965).
- <sup>28</sup> J. v. Euw and T. Reichstein. *Helv. Chim. Acta* **47**, 711 (1964).
- <sup>29</sup> J. W. Cornforth, R. H. Cornforth, A. Pelter, M. G. Horning, and G. Popják. *Tetrahedron* **5**, 311 (1959).
- <sup>30</sup> N. Neuss and M. Gorman. *Tetrahedron Letters* 206 (1961).
- <sup>31</sup> G. F. Smith and M. A. Wahid. *J. Chem. Soc.* 4002 (1963).
- <sup>32</sup> H. Conroy, P. R. Brook, M. K. Rout, and N. Silverman. *J. Am. Chem. Soc.* **80**, 5178 (1958); H. Conroy, P. R. Brook and Y. Amiel. *Tetrahedron Letters* No. 11, 4 (1959).
- <sup>33</sup> D. A. Yeowell and H. Schmid. *Experientia* **20**, 250 (1964).
- <sup>34</sup> A. Mondon and G. Teege. *Chem. Ber.* **91**, 1014 (1958).
- <sup>35</sup> A. R. Battersby, R. T. Brown, J. A. Knight, J. A. Martin and A. O. Plunkett. *Chem. Comm.* 346 (1966).
- <sup>36</sup> P. Loew, H. Goeggel, and D. Arigoni. *Chem. Comm.* 347 (1966).
- <sup>37</sup> E. S. Hall, F. McCapra, T. Money, K. Fukumoto, J. R. Hanson, B. S. Mootoo, G. T. Phillips, and A. I. Scott. *Chem. Comm.* 348 (1966).
- <sup>38</sup> A. R. Battersby and B. J. T. Harper. *J. Chem. Soc.* 1748 (1959).
- <sup>39</sup> G. Büchi and R. E. Manning. *Tetrahedron* **18** 1049 (1962).
- <sup>40</sup> C. Djerasi, T. Nakano, A. N. James, L. H. Zalkow, E. J. Eisenbraun, and J. N. Shoolery. *J. Org. Chem.* **26**, 1192 (1961).
- <sup>41</sup> H. Inouye, T. Arai, and Y. Miyoshi. *Chem. Pharm. Bull. Japan*, **12**, 888 (1964).
- <sup>42</sup> A. J. Birch and J. Grimshaw. *J. Chem. Soc.* 1407 (1961); K. Sheth, E. Ramstad, and J. Wolinsky. *Tetrahedron Letters* 394 (1961), and references therein.

A. R. BATTERSBY

- <sup>43</sup> A. R. Battersby and J. J. Reynolds. *Proc. Chem. Soc.* 346 (1960).  
<sup>44</sup> E. Leete and P. E. Nemeth. *J. Am. Chem. Soc.* **82**, 6055 (1960);  
E. Leete. *J. Am. Chem. Soc.* **85**, 3666 (1963).  
<sup>45</sup> A. R. Battersby, R. Binks, J. J. Reynolds, and D. A. Yeowell. *J. Chem. Soc.* 4257 (1964).  
<sup>46</sup> A. R. Battersby, R. Binks, and D. A. Yeowell. *Proc. Chem. Soc.* 86 (1964);  
A. R. Battersby and R. B. Herbert. *Proc. Chem. Soc.* 260 (1964).  
<sup>47</sup> cf. O. L. Chapman and P. Fitton. *J. Am. Chem. Soc.* **85**, 41 (1963).  
<sup>48</sup> E. Leete. *Tetrahedron Letters* 333 (1965).  
<sup>49</sup> J. Hrbek, jun., and F. Šantavý. *Coll. Czech. Chem. Comm.*, **27**, 255 (1962).  
<sup>50</sup> A. R. Battersby, R. B. Herbert, L. Pijewska, and F. Šantavý. *Chem. Comm.* 228 (1965).  
<sup>51</sup> M. P. Cava, K. Nomura, R. H. Schlessinger, K. T. Buch, B. Douglas, R. F. Rattauf, and  
J. A. Weisbach. *Chem. and Ind.* 222 (1964).  
<sup>52</sup> A. R. Battersby, M. Hirst, and J. Staunton. Unpublished work.

# **BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS**

F. LYNEN

*Max-Planck-Institute for Cell Chemistry, Münich, Germany*

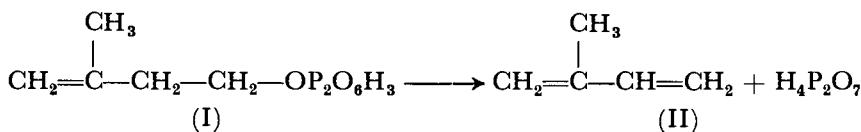
## **INTRODUCTION**

Investigations during the last three decades have shown that acetic acid is one of the most important building units for the synthesis of complex organic molecules found in Nature. This insight is primarily due to the application of tracer techniques to biosynthetic problems. Despite all the admiration this method deserves, we should not forget, however, that organic chemists, full of imagination, earlier recognized common structural features in different molecules, and postulated biosynthetic schemes and special rules, which could be used in structural studies on new natural products with great success. The special rules I have in mind are the "isoprene rule" of Ruzicka<sup>1</sup> and the "polyacetate rule" of Birch<sup>2</sup>. Today every organic chemist interested in natural products must be familiar with both rules, and thus I need not dwell on their wide-spread application. Furthermore, a number of excellent reviews on this subject exist (cf. ref. 3). Therefore I want to restrict myself to indicating briefly the fundamental principles of both rules, and then discuss in greater detail the enzymatic systems which participate in the accompanying biosynthetic reactions. Here I want to emphasize some recent studies in my laboratory.

## **BIOSYNTHETIC PATHWAY via "ISOPRENE RULE"**

The "isoprene rule" originated from structural investigations of the volatile terpene oils. Otto Wallach and others studying the structural class of mono-terpenes realized the relationship between this group of compounds and the simple unsaturated hydrocarbon isoprene. But it was Ruzicka's fundamental observation that this structural element is broadly distributed among many classes of natural products<sup>1</sup>.

The "isoprene rule" was clarified when  $\Delta^3$ -isopentenyl pyrophosphate (I) was identified to be the building block of the terpenes<sup>4,5</sup>. Isopentenyl pyrophosphate can be considered as a "disguised" isoprene because elimination of pyrophosphoric acid would lead to isoprene (II).



## Terpene biosynthesis

### Biosynthesis involving isopentenyl pyrophosphate—the later phase

The synthesis of the terpenoid carbon chains involves C-alkylation (cf. Figure 1). This is made possible by the bifunctional nature of isopentenyl pyrophosphate, based on the nucleophilic reactivity of the double bond and the potential electrophilic character of the pyrophosphate ester. The process is initiated by the isomerization of isopentenyl pyrophosphate to  $\gamma,\gamma$ -dimethylallyl pyrophosphate by a shift of the double bond. With the formation of dimethylallyl pyrophosphate the electrophilic reactivity of the precursor isopentenyl pyrophosphate is fully realized. Ionization of the carbon–oxygen bond of the dimethylallyl pyrophosphate creates a cationic centre

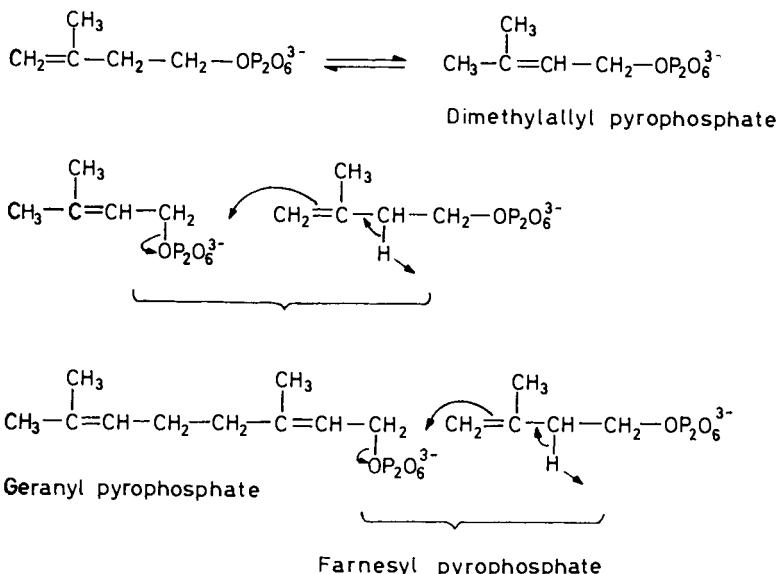


Figure 1. The mechanism of the synthesis of isoprenoid carbon chains

which then attacks the electrons available in the exomethylene group of isopentenyl pyrophosphate. Subsequently, the elimination of a proton leads to the first condensation product, geranyl pyrophosphate. This homologous allyl pyrophosphate by acquisition of another isopentenyl pyrophosphate is converted into farnesyl pyrophosphate, the direct precursor of the sesquiterpenes. A continuation of this process leads to geranyl-geranyl pyrophosphate, from which the diterpenes can be derived, or to the higher linear arrays that are found in the ubiquinones, plastoquinones or gutta percha. The steric course of chain growth, whether olefinic linkages with *trans* or *cis* stereochemistry are produced, depends on the stereospecificity of the enzyme involved. As was demonstrated by Archer *et al.*<sup>6</sup>, the polymerizing enzyme from Hevea latex, catalysing the production of natural rubber (the all *cis* polyisoprene), eliminates a proton epimeric with that eliminated in the synthesis of the *trans* compound farnesyl pyrophosphate.

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

The reactivity of allylic pyrophosphate may be due to two combined effects: (*i*) the ease of displacement of the pyrophosphate anion, and (*ii*) the resonance stabilization of the allylic carbonium ion. The fact that pyrophosphoric acid is more strongly acidic than orthophosphoric acid and consequently is a better leaving group is also evidenced by the increased rate of acid hydrolysis of allyl pyrophosphates compared with that of allyl phosphates<sup>7</sup>. In addition, the pyrophosphate residue of isopentenyl pyrophosphate may exert a neighbouring group effect on the alkylation, as suggested by Johnson and Bell<sup>8</sup>.

The crucial role played by the isomerase in the chain of terpene biosynthesis is obvious. It catalyses the isomerization of isopentenyl pyrophosphate to the allylic derivative and thus triggers the synthesis of the terpene chain. Once this process has started it can continue without renewed isomerization, and is limited only by the specificity of the participating enzymes. This was first shown in our laboratory by experiments with the enzyme system of yeast catalysing the synthesis of farnesyl pyrophosphate<sup>9</sup>. We were able to demonstrate it again in more recent studies on rubber biosynthesis.

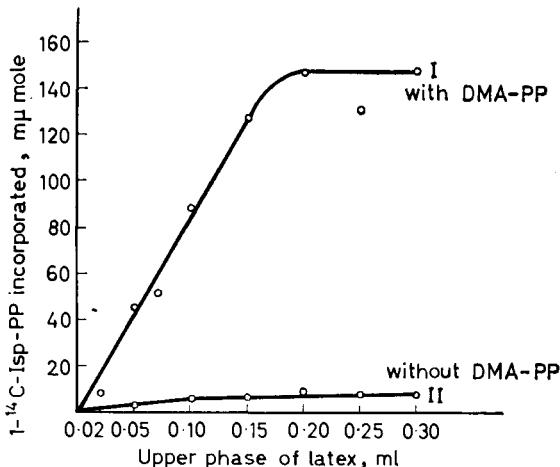
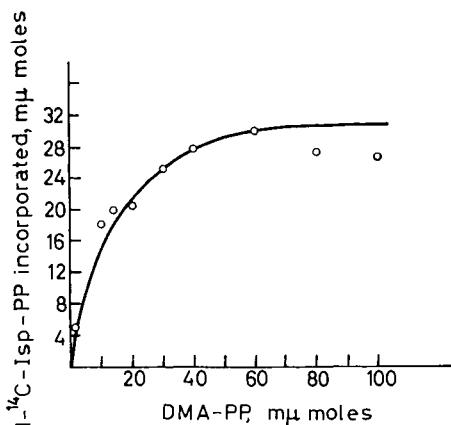


Figure 2. Incorporation of  $1-^{14}\text{C}$ -isopentenyl pyrophosphate (Isp-PP) into rubber with and without addition of dimethylallyl pyrophosphate (DMA-PP). An 0.8 ml reaction mixture contained (in  $\mu\text{moles}$ ): TRIS buffer, pH 8.0, 100;  $\text{MgK}_2$ -ethylenediamine tetraacetate, 20;  $\text{MgCl}_2$ , 2; cysteine, 5;  $1-^{14}\text{C}$ -isopentenyl pyrophosphate ( $2.1 \times 10^5$  c.p.m./ $\mu\text{mole}$ ), 0.17; and the resuspended creamy upper phase of centrifuged *Hevea* latex, as indicated on the abscissa. The experiments of curve I contained in addition 0.1  $\mu\text{moles}$  of dimethylallyl pyrophosphate. After 120 min incubation at  $26^\circ\text{C}$  the rubber was isolated and its radioactivity measured.

As Henning<sup>10</sup> had found in my laboratory some years ago, freshly tapped latex from the rubber tree *Hevea brasiliensis* catalyses the incorporation of synthetic  $1-^{14}\text{C}$ -isopentenyl pyrophosphate into rubber, a fact which was later confirmed by Archer *et al.*<sup>11</sup> In collaboration with Hopper, Nordwig, Berndt, and Dick we continued these studies. One of the problems involved was to find a good supply of *Hevea* latex. By the generous support of Professor Milanez, then director of the famous botanical garden in Rio de Janeiro, we were allowed to use one of the *Hevea* trees of the garden as a

F. LYNEN

source of latex. Dr. Raoul Machado expertly tapped the tree, collected the latex, added cysteine and bicarbonate as protective agents and sent the material, after freezing it in dry ice, by aeroplane to Munich. In this way we received the enzymatically active material 24 hours after its harvest in Rio de Janeiro. By centrifugation of the thawed material at 20 000 g it separated into a slightly opalescent aqueous bottom phase, and a creamy top layer, which contained the bulk of the preformed rubber and the polymerizing enzyme. This polymerase remained in the creamy layer during repeated washing with TRIS-buffer, pH 8.0, containing serum albumin and TWEEN. By this treatment the isomerase was removed completely, and the incorporation of  $1^{14}\text{C}$ -isopentenyl pyrophosphate into high molecular rubber became now strictly dependent on the addition of dimethylallyl pyrophosphate as "primer" of the polymerization (*Figure 2*). The Michaelis constant of the polymerase for dimethylallyl pyrophosphate was found to be  $1.3 \times 10^{-5} \text{ M}$  (*Figure 3*).



*Figure 3.* Dependence of rubber synthesis on the concentration of dimethylallyl pyrophosphate. Experimental conditions were identical to those of *Figure 2* except that the final volume was 0.5 ml. The reaction mixtures contained 0.1 ml of upper phase of Hevea latex and dimethylallyl pyrophosphate (DMA-PP) as indicated on the abscissa

These results may be interpreted with the assumption that the polymerization of  $\text{C}_5$ -units to high molecular rubber is catalysed by one enzyme. The close association of this enzyme with the rubber droplets of the emulsion appears most appropriate from the physiological point of view, because otherwise the reaction product, rubber, could not be removed from the surface of the enzyme. We assume therefore that the polymerase is located at the surface area of the small rubber droplets, where it can accept the water-soluble substrates from the serum and deliver the synthesized hydrocarbon into the hydrophobic phase. In contrast to the polymerase the enzymes involved in the biosynthesis of isopentenyl pyrophosphate from acetate, which will be discussed later, are present in the aqueous serum fraction. Anticipating any discussion, I would like to present in *Table 1* the

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

Table 1. Activity of the enzymes involved in rubber synthesis

<i>Enzyme</i>	<i>Enzyme activity (37°)</i> (mμmoles of substrate/min/ml latex)
Acetyl-CoA synthetase	59
Thiolase	3.92 (3920†)
HMG-CoA synthase	232
HMG-CoA reductase	0.078
Mevalonate kinase	149
P-Mevalonate kinase	44
PP-Mevalonate decarboxylase	103
Isopentenyl-PP isomerase	—
Polymerase	22.3

† for the reverse reaction.

results of our quantitative measurements of the various enzyme activities in the serum fraction. As can be seen all enzyme activities are of comparable magnitude with the exception of the  $\beta$ -hydroxy- $\beta$ -methylglutaryl(HMG)-CoA reductase. This enzyme activity, which is responsible for the formation of mevalonic acid, is much lower, and it is conceivable that this enzymatic step may represent the physiological bottleneck of rubber biosynthesis. From the point of view of comparative biochemistry this observation seems to be very important since HMG-CoA reduction is also the rate limiting step of cholesterol synthesis in the mammalian organism (cf. page 148).

There is every reason to believe that isopentenyl pyrophosphate functions universally as the "monomeric" precursor for the great variety of linear and cyclic isoprene derivatives which occur in Nature. The sequence of events seems to be always the same. Following the synthesis of open chain terpenyl derivatives, secondary transformations such as reduction, cyclization, rearrangement, and oxidation reactions lead to the reduced, cyclized, and oxygenated terpenoid compounds. Interception by an activated aromatic ring or by protohaeme will lead to natural products like ubiquinones, vitamins K or haeme a<sup>12</sup>.

I do not want to go further into details about secondary transformations of open chain terpenyl derivatives, because apart from the synthesis of lanosterol and cholesterol, little is known about the enzymes involved. Biosynthetic experiments with isotopically labelled precursors have in many cases revealed that reaction mechanisms predicted by applying the rules of organic chemistry agree very well with the isotope distribution found experimentally (cf. ref. 3). The most brilliant example is the experimental verification of the theoretical model of triterpene biogenesis, constructed by the Zürich school<sup>13</sup>.

Among the various transformations in the field of terpene biochemistry, the dimerization of farnesyl pyrophosphate to squalene deserves particular mention. This process (cf. Figure 4) represents the prototype of the "tail to tail" linkage of C<sub>5</sub>-units, first represented by Ruzicka<sup>1</sup> as an irregular type of condensation in the biosynthesis of terpenes. Since the reductive dimerization of two molecules of farnesyl pyrophosphate occurs with simultaneous elimination of the pyrophosphate residues, this reaction in a formal sense

F. LYNEN

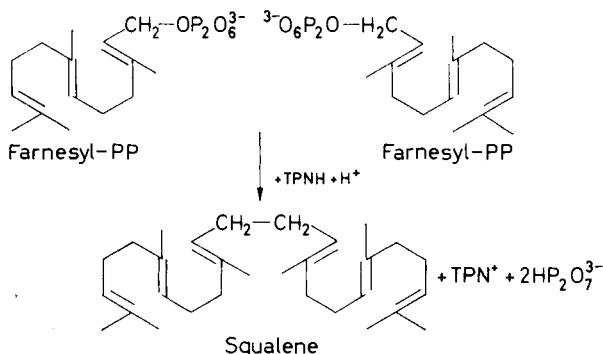


Figure 4. Balance equation for the formation of squalene from farnesyl pyrophosphate

resembles Karrer's<sup>14</sup> chemical synthesis of squalene from farnesyl bromide and magnesium.

In the case of the enzymatic synthesis TPNH assumes the role of electron donor. The detailed chemical mechanism of squalene synthesis from two molecules of farnesyl pyrophosphate is still obscure<sup>3</sup>. But the beautiful work of Cornforth *et al.*<sup>15</sup> has not only elucidated the stereochemistry of this process but also presented unequivocal proof that the condensation is an asymmetric process in the sense that one of the farnesyl pyrophosphate molecules condensing is subject to reactions different from those of the other.

#### Biosynthesis involving mevalonic acid—the earlier phases

Thus far I have discussed the later phase of terpene biosynthesis. Let us turn now to the discussion of earlier phases in which mevalonic acid occupies a central position. The discovery of mevalonic acid by Wright, Folkers, and associates (cf. ref. 16) at the Merck, Sharp, and Dohme laboratories ten years ago proved to be the real starting point for the elucidation of the chemical mechanism of terpene biosynthesis. Enzymatic experiments with radioactively labelled mevalonic acid in cell-free extracts and with purified enzymes from yeast or liver led to the isolation of "active isoprene" and its identification as isopentenyl pyrophosphate (cf. ref. 3).

The biological precursor of mevalonic acid (III) is  $\beta$ -hydroxy- $\beta$ -methyl-glutaryl-CoA (IV), which is enzymatically reduced by TPNH (Figure 5). The particular enzyme involved in this reduction was purified from yeast extracts<sup>17,18</sup>. When its mode of action was studied in greater detail it was found that free mevaldic acid (V) is not an intermediate in this reaction. For example, added mevaldic acid was only very slowly reduced by the purified enzyme. According to recent experiments of Dr. Rétey<sup>19</sup>, the semi-mercaptal of mevaldic acid with coenzyme A (VI) is much more rapidly reduced by the purified enzyme from yeast and seems to satisfy the kinetic requirements of an intermediate.

Pertaining to the conversion of mevalonic acid (III) to  $\Delta^3$ -isopentenyl pyrophosphate (I; cf. Figure 6) by way of 5-phospho- (VII) and 5-pyro-phosphomevalonic acid (VIII), the step that generates the biological

BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

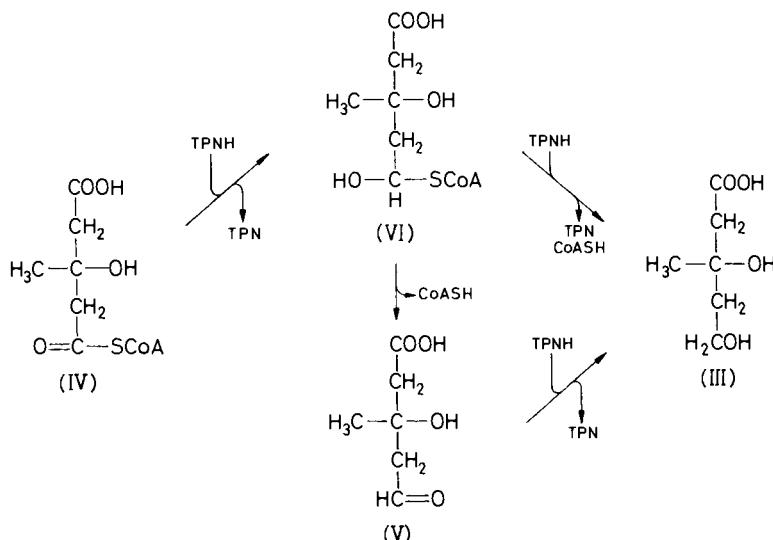


Figure 5. Possible mechanism of mevalonic acid formation from  $\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA

isoprene unit is chemically the most interesting reaction. The enzyme involved catalyses the coordinated removal of the carboxyl group and of the *tertiary* hydroxyl function. Data of Bloch *et al.*<sup>20</sup> obtained with  $^{18}\text{O}$  suggest that 3-phosphomevalonic-5-pyrophosphate (IX), is a transitory intermediate, ATP serving as the phosphorylation agent for the *tertiary* hydroxyl group thereby promoting its elimination.

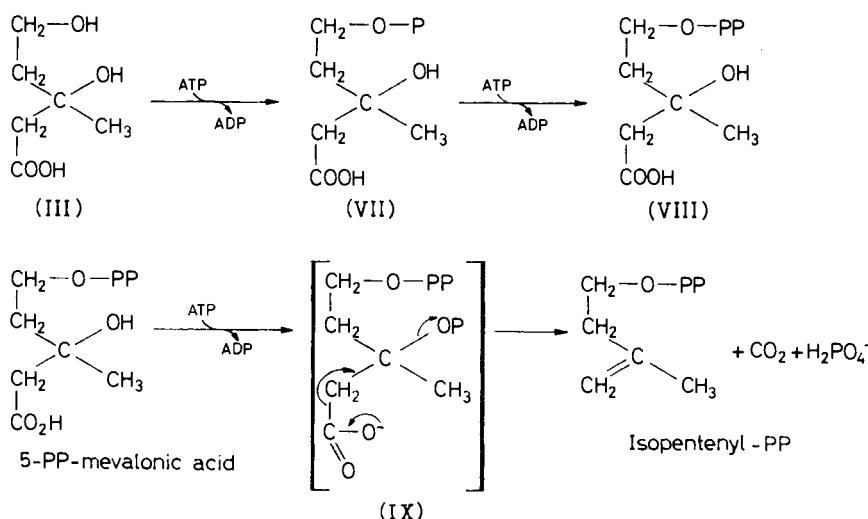


Figure 6. The formation of  $\Delta^3$ -isopentenyl pyrophosphate from mevalonic acid

## F. LYNEN

Especially relevant to the topic of my lecture are the reactions leading to  $\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA from acetyl-CoA. Here the carbon skeleton of the C<sub>6</sub>-precursors of isopentenyl pyrophosphate is assembled by carbon to carbon condensation of three acetate units. The complete understanding of these reactions depends on the realization that an acetyl thioester is activated both at the carbonyl and the methyl group ("head" and "tail" end)<sup>21</sup>. In the condensation of two acetyl-CoA to form acetoacetyl-CoA both reactivities of the thioester are involved (Figure 7). One molecule

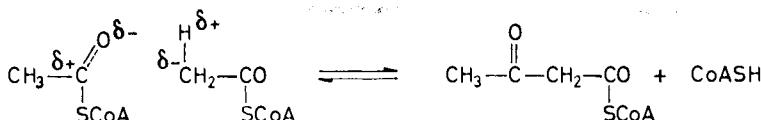


Figure 7. The mechanism of acetoacetyl-CoA formation

of acetyl-CoA is the nucleophilic agent and acetate acceptor and the other is the electrophilic acetyl donor. The condensation reaction is reversible and under physiological conditions the formation of acetyl-CoA from acetoacetyl-CoA is by far the preferred reaction. By means of spectrophotometric techniques it was possible to establish that at pH 7 the equilibrium is<sup>21</sup>:

$$K_{\text{eq}} = \frac{[\text{acetoacetyl-CoA}] \times [\text{CoA}]}{[\text{acetyl-CoA}]^2} = 1.6 \times 10^{-5}$$

Nevertheless, it is possible to achieve the synthesis of carbon chains by thiolase reaction under physiological conditions, if the acetoacetyl-CoA produced is removed by subsequent condensation with acetyl-CoA (Figure 8). The condensation of acetyl-CoA and acetoacetyl-CoA to form  $\beta$ -hydroxy- $\beta$ -methyl-glutaryl-CoA and free CoA seems to be irreversible. At least it was not

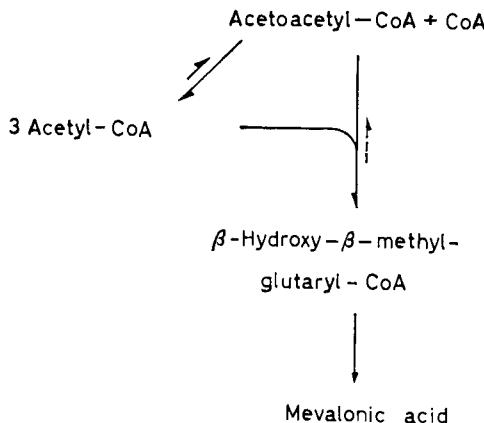


Figure 8. The coupling of thiolase reaction with hydroxymethylglutaryl-CoA synthase reaction

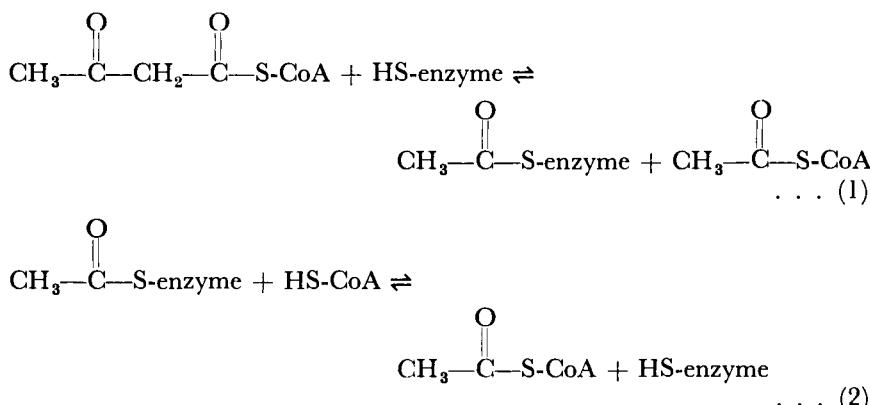
## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

possible to demonstrate any reversal of this condensation reaction even under conditions in which this should have been favoured<sup>17,18</sup>. Thus by coupling with the condensation the thiolase reaction can be forced in the direction of acetoacetyl-CoA formation with subsequent accumulation of  $\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA.

When I first studied the properties of the enzyme thiolase I found its high sensitivity against sulphhydryl blocking agents and postulated that the active centre contained a sulphhydryl group, which participated in the catalytic process<sup>22</sup>. According to this hypothesis the acetoacetyl-CoA is cleaved by the SH-enzyme with the generation of acetyl-CoA and an acetyl-S-enzyme. The acetyl-S-enzyme then transfers its acyl radical to coenzyme A with the regeneration of the free SH-enzyme. Therefore thiolase should combine two activities:

Thiolytic activity as illustrated by reaction (1).

Acylytransferase activity as illustrated by reaction (2).



Gehring in my laboratory has now finally been able to prove this two step mechanism<sup>23</sup>. He has achieved a 700-fold purification of thiolase beginning with crude extracts from pig heart. The purified enzyme was homogeneous in the ultracentrifuge and could be crystallized by fractionation with ammonium sulphate. Its molecular weight was found<sup>24</sup> to be 168 000.

In *Figure 9* is summarized an experiment in which the reversible thiolytic step (a) was demonstrated independently of the acyltransferase step (b). Acetoacetyl-CoA was incubated with the purified enzyme and 1-<sup>14</sup>C-acetyl-CoA for various periods of time and the distribution of radioactive carbon between carbonyl and carboxyl group of the coenzyme A bound acetoacetate was determined. As can be seen in *Figure 9*, the carboxyl carbon became highly radioactive whereas the carbonyl carbon remained practically without radioactivity. Similar asymmetric labelling was obtained previously by Beinert and Stansly<sup>25</sup> with crude enzyme preparations. An interpretation of the asymmetric labelling of acetoacetate is as follows: The nonlabelled acetoacetyl-CoA forms nonlabelled acetyl enzyme in the forward reaction (1). This acetyl enzyme now condenses with 1-<sup>14</sup>C-labelled acetyl-CoA in

F. LYNEN

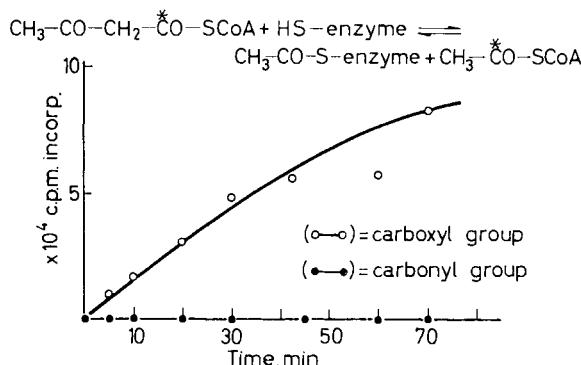


Figure 9. Incorporation of  $1\text{-}^{14}\text{C}$ -acetyl-CoA into the carboxyl group of acetoacetyl-CoA with thiolase<sup>23</sup>. 2 ml reaction mixture, kept at  $0^\circ\text{C}$ , contained in  $\mu\text{moles}$ : K-phosphate, pH 7.5, 60; acetoacetyl-CoA, 0.5;  $1\text{-}^{14}\text{C}$ -acetyl-CoA, 0.3 ( $3.35 \times 10^6$  c.p.m.); 3 mg serum albumin and 50  $\mu\text{g}$  thiolase (sp. activity 6). After various times of incubation aliquots were taken and the isotope distribution between keto carbon and carboxyl carbon was determined

the backward reaction (1) to give carboxyl-labelled acetoacetyl-CoA. The rate of interchange of acetyl groups between acetyl-CoA and acetyl enzyme is negligible because of the lack of free coenzyme A.

The formation of acetyl enzyme is further supported by the following experiment (Figure 10). Varying amounts of the purified enzyme were preincubated with  $1\text{-}^{14}\text{C}$ -labelled acetyl-CoA, then precipitated with trichloroacetic acid and carefully washed. As can be seen in Figure 10 the protein became radioactive in proportion to the amount of protein added, as to be expected if acetyl enzyme was generated with the acetyl radical covalently bound to the protein (equation 2). Studies on the interaction of thiolase with  $1\text{-}^{14}\text{C}$ -labelled iodoacetamide finally lead to the identification

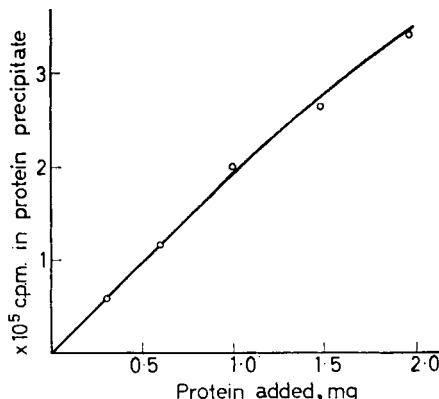
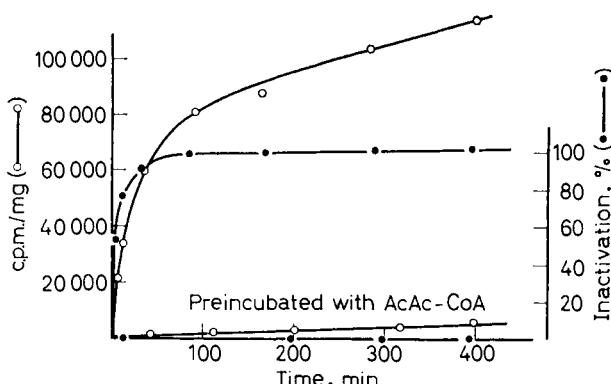


Figure 10. Demonstration of acetyl enzyme formation<sup>23</sup>. Each reaction mixture, vol. 1 ml, contained (in  $\mu\text{moles}$ ): K-phosphate, pH 7.0, 15; glutathione, 5;  $1\text{-}^{14}\text{C}$ -acetyl-CoA, 0.1 ( $2.7 \times 10^6$  c.p.m.); 4 mg of serum albumin and thiolase, as indicated on the abscissa. After incubation for 10 min at  $10^\circ\text{C}$ , the protein was precipitated with trichloroacetic acid, carefully washed and its fixed radioactivity determined

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

of the acceptor group on the enzyme. In these experiments (cf. *Figure 11*) the kinetics of the inactivation of thiolase by  $1\text{-}^{14}\text{C}$ -iodoacetamide (full circles) was compared with the amount of iodoacetamide bound to the protein (open circles), which was measured by the radioactivity precipitated with trichloroacetic acid. As seen from *Figure 11* the free enzyme is alkylated by iodoacetamide and is inactivated at the same time, whereas the acetyl enzyme generated by preincubation of the enzyme with acetoacetyl CoA (equation 1) is protected against interaction with iodoacetamide, and therefore remains enzymatically active. Thus it is reasonable to



*Figure 11.* Interaction of  $1\text{-}^{14}\text{C}$ -iodoacetamide with thiolase. Two reaction mixtures, vol. 5 ml, each contained 0.02 M TRIS-buffer, pH 8.2;  $5 \times 10^{-5}$  M  $1\text{-}^{14}\text{C}$ -iodoacetamide and 5.2 mg thiolase. One reaction mixture was preincubated with  $7 \times 10^{-4}$  M acetoacetyl-CoA before iodoacetamide was added. After various incubation times at 0°C, as indicated on the abscissa, aliquots were withdrawn and thiolase activity (●—●) as well as protein bound radioactivity (○—○) were measured

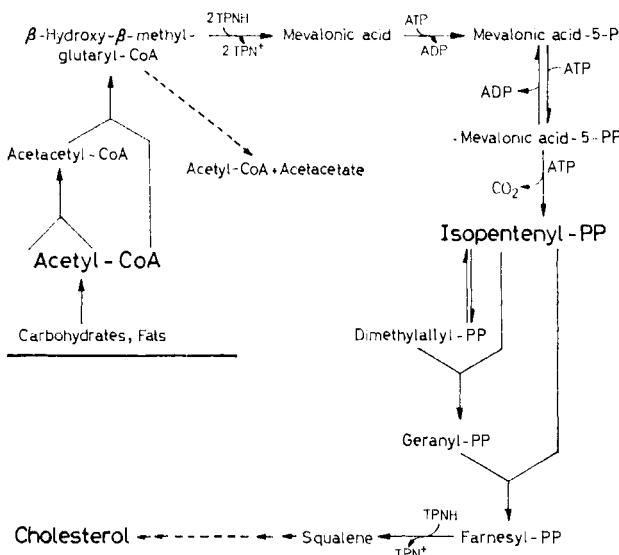
assume that iodoacetamide reacts specifically with the acetyl acceptor group of the enzyme. Upon cleavage of the iodoacetamide treated enzyme with hydrochloric acid the radioactive fission product was identified as S-carboxymethyl cysteine by comparison with the authentic compound<sup>23</sup>. Quantitative measurements of the acetyl- and iodoacetamide binding capacity revealed that 3.6–3.7 residues are bound per enzyme molecule of molecular weight 168 000, which would indicate that thiolase belongs to the group of enzymes with tetrameric structure and is composed of four subunits each carrying one active SH-group. This assumption is further supported by studies on the reversible dissociation and inactivation of the enzyme treated with 5 M urea<sup>24</sup>.

It has been postulated<sup>26</sup> that the initial condensation leading to the acetoacetate-level in terpene synthesis involves an acetyl unit and a malonyl unit as in the synthesis of long chain fatty acids. Energetically the mechanism would certainly be more favourable than the thiolase reaction. Bloomfield and Bloch<sup>27</sup> have observed, however, that in biotin deficient yeast sterol synthesis from acetate proceeds normally, whereas fatty acid synthesis is greatly impaired. This result would seem to argue against any role of biotin catalysis, and therefore of malonyl units, in terpene synthesis from acetate.

The reaction sequence leading to terpenyl pyrophosphates, the precursors

F. LYNEN

for all naturally occurring linear and cyclic isoprene derivatives, is summarized in *Figure 12*. The sequence can be divided into three sections. In the first section, the carbon skeleton of mevalonic acid is synthesized from three molecules of acetyl-CoA. The energy required for this process is derived from the three thioester bonds, which are cleaved to regenerate free coenzyme A, and the TPNH molecules, which become oxidized. In the second section



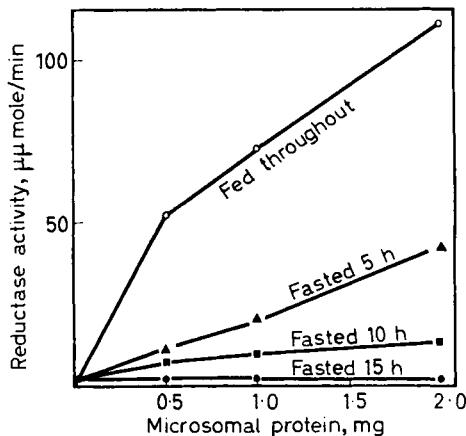
*Figure 12.* The route of the biosynthesis of polyterpenes

the conversion of mevalonic acid to isopentenyl pyrophosphate so suitable as the building unit for many natural products is achieved. This activation requires three moles of ATP and is simultaneously driven by a decarboxylation reaction. For the synthesis of the polyisoprenoid carbon chains in the last section, no additional source of energy is required, sufficient being made available by the expulsion of inorganic pyrophosphate from the allyl derivatives and its subsequent hydrolysis to orthophosphate by pyrophosphatase which is ubiquitous to all living systems.

All evidence to date indicates that changes in the activity of the enzyme responsible for the reduction of hydroxymethylglutaryl-CoA to mevalonic acid are in a direction and of a magnitude to account for the physiological control of polyisoprenoid biosynthesis. This was first realized in studies on cholesterol synthesis<sup>28</sup> which can be virtually abolished by cholesterol feeding<sup>29-31</sup> and by fasting<sup>32,33</sup>, or is increased by x-irradiation<sup>34</sup> and intravenous injection of Triton WR-1339<sup>35</sup>. The inhibition of hydroxymethylglutaryl-CoA reductase by cholesterol feeding seems to indicate that the principle of negative feed back operates in sterol biosynthesis as it does in so many pathways. With these considerations in mind the reductase step seems ideally suited to the control function, since it is the first step on the direct pathway to the polyisoprenoids and is essentially irreversible. Further-

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

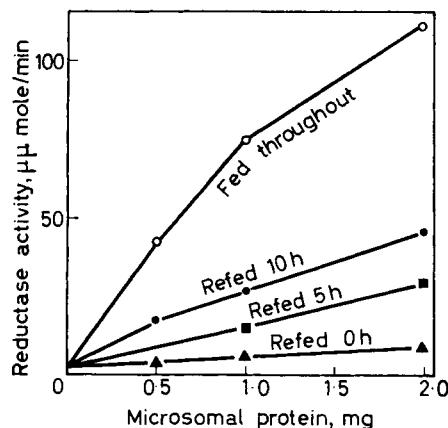
more hydroxymethylglutaryl-CoA generated in the condensation and not reduced to mevalonic acid, by the interaction of hydroxymethylglutaryl-CoA cleavage enzyme can be split into acetyl-CoA and acetoacetate, and thus channelled back to the acetate pool. Resuming an investigation which



*Figure 13.* Effect of fasting on HMG-CoA-reductase activity in rat liver microsomes<sup>37</sup>

was started by Bucher<sup>36</sup> during a short visit to my laboratory, Regen *et al.*<sup>37</sup> systematically studied the changes in hydroxymethylglutaryl-CoA reductase activity of rat liver homogenates during a period of starvation and refeeding of animals.

The method for estimating the enzyme involved preparation of <sup>14</sup>C-labelled hydroxymethylglutaryl-CoA and the separation of the labelled reduction product, mevalonic acid, by Celite column chromatography. The reductase was found to be firmly attached to the membranous elements of the microsome fraction and to undergo drastic changes incident on brief



*Figure 14.* Effects of refeeding after a 20 h fast on the HMG-CoA reductase of liver microsomes<sup>37</sup>

## F. LYNEN

food deprivation. We were surprised by the promptness of the response as seen in *Figure 13*. After only 5 hours the activity was less than 50 per cent of the fed control. The sudden fall in activity could represent a diminution of enzyme protein or a decrease in the catalytic ability of the protein due to a firmly bound inhibitor. At present further experiments are in progress which might decide between both possibilities. The effect of the standard diet on this enzyme is further illustrated by the restoration of activity upon refeeding after a 20-hour fast (cf. *Figure 14*). The response in this case appears somewhat slower.

### BIOSYNTHETIC PATHWAY via "POLYACETATE RULE"

I shall now focus attention on the second pathway from acetate to natural products, which is governed by the "polyacetate rule". This biogenetic concept was first realized by Collie<sup>38</sup> on the basis of some interesting transformations he had observed with polyacetyl compounds. However, it was not until 1953, when Birch<sup>2</sup> in a brilliant re-examination of the concept laid down most of the modern view of the polyacetate hypothesis and not only pointed out its broad scope and great usefulness but also accumulated experimental support by studies with radioactive tracers.

The primary building block of this biosynthetic pathway was assumed to be a linear poly- $\beta$ -ketomethylene, formed by repeated head-to-tail condensation

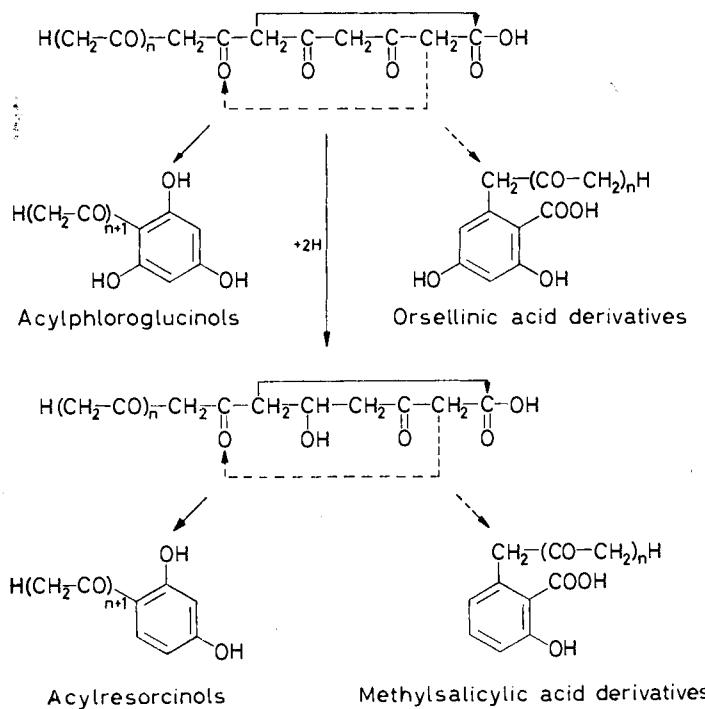


Figure 15. Illustration of the polyacetate-rule

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

of acetate units (*Figure 15*). By secondary transformations, such as reduction, dehydration, condensation, cyclization, rearrangement, and decarboxylation reactions polyacetate molecules of different chain length can lead to an immense variety of naturally occurring organic molecules. Especially since additional variants are given by the introduction of methyl or isopentenyl groups at the methylene sites, or by replacement of acetic acid in the terminal position by another acid such as linear or branch chain aliphatic acids or some benzoic and cinnamic acids<sup>3</sup>.

One of the key reactions of the polyacetyl chain is cyclization. Thus an enolate anion may attack the carbonyl carbon of either keto or carboxyl group at another position of the chain to yield cyclized products, as illustrated in *Figure 15*. Both reactions occur with facility in forming six-membered rings and lead to two families of phenols, the acylphloroglucinol and orsellinic acid derivatives, respectively, distinguished by the aromatic substituent pattern. By reduction of a carbonyl group not involved in the cyclization and ring-closure by loss of water derivatives of methylsalicylic acid or acyl resorcinols respectively are generated.

With the detection of the thioester bond in acetyl-CoA and the identification of the enzyme, thiolase, the enzymatic problem of the polyacetate synthesis seemed to be essentially solved<sup>39</sup>. It was assumed that repeated condensations of acetyl-CoA molecules could generate poly- $\beta$ -ketomethylene structures. On the other hand the value found for the equilibrium of the thiolase reaction, mentioned earlier in my lecture, was difficult to reconcile with this hypothesis. By simple calculations we were able to demonstrate that even the synthesis of 3,5-diketohexanoyl-CoA by a thiolase type condensation of three acetyl-CoA molecules was thermodynamically impossible under physiological conditions<sup>40</sup>.

### Fatty acid synthesis

This difficulty was overcome when studies on fatty acid synthesis revealed malonyl-CoA as the building block for carbon-chain synthesis. In the condensation with malonyl-CoA the equilibrium is shifted to the side of the  $\beta$ -keto compound by the accompanying decarboxylation.

With these considerations in mind I postulated at the Gatlinburg conference in 1959 that malonyl-CoA is the precursor of all natural products with polyacetate structure<sup>41</sup>. In the following years this hypothesis has found experimental support by the investigations of many laboratories<sup>3</sup>. In these experiments, after feeding isotopically labelled precursors to living systems, the distribution of isotopic atoms in the synthesized products was determined by controlled chemical degradation and was found to be in agreement with the predictions. However, enzymatic studies in this field are rare and are mainly restricted to the exploration of fatty acid synthesis.

Before discussing the enzyme system of fatty acid synthesis I would like to dwell for a few moments on the enzyme responsible for the generation of malonyl-CoA by carboxylation of acetyl-CoA. This reaction belongs to the class of carboxylation processes, in which biotin participates as the coenzyme and acts as a CO<sub>2</sub>-carrier in the overall process (*Figure 16*). As elucidated by our studies<sup>12</sup>, biotin, covalently bonded to the protein, is first carboxylated

### F. LYNEN

by bicarbonate at the expense of ATP which is split into ADP and inorganic phosphate. The carboxybiotin enzyme thus formed then transfers its

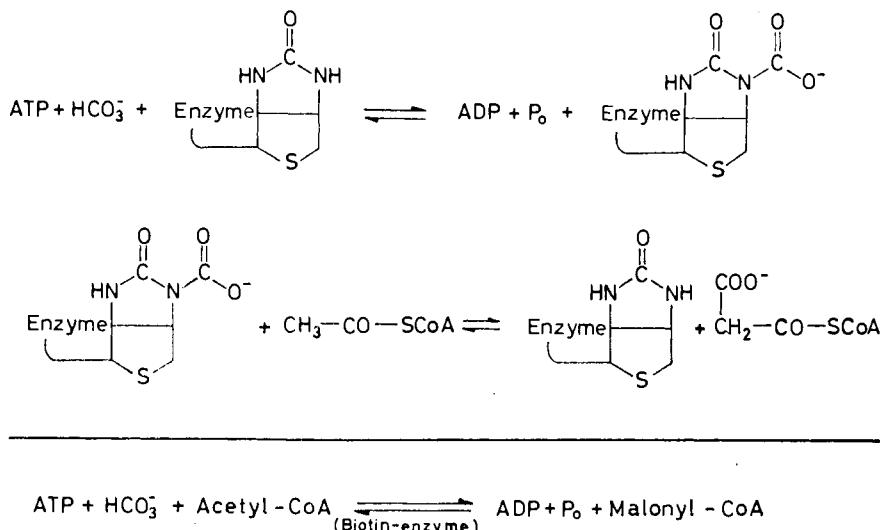


Figure 16. Mechanism of the carboxylation of acetyl-CoA

carboxyl group to acetyl-CoA yielding malonyl-CoA. In the carboxylated enzyme 1'-*N*-carboxybiotin is bound to the enzyme protein through amide linkage at the  $\epsilon$ -amino group of a lysine residue (cf. Figure 16). The chemical reactivity of an "active carbonic acid" of this structure derives from the electron attraction of the ureido system. The bond between biotin and carbon dioxide is polarized towards nitrogen which augments the electrophilic character of the carboxyl group and its reaction with the enolate anion of acetyl-CoA. The process was found to be reversible and may occur by way of a concerted cyclic process as depicted in Figure 17.

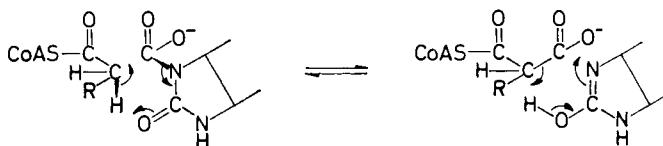


Figure 17. Chemical mechanism of transcarboxylation

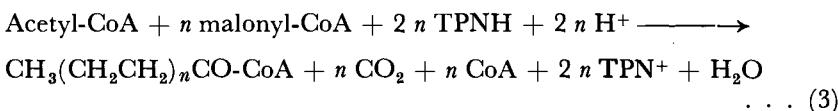
#### *Enzyme system of fatty acid synthesis*

Insight into the chemical details of fatty acid synthesis, starting with malonyl-CoA, was first obtained in studies with the enzyme system of yeast<sup>40,42</sup>. From yeast cells, ruptured by vigorous shaking with glass beads, we were

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

able to isolate a protein fraction 150 times more active than the crude extract. The purified enzyme, which we named fatty acid synthetase, proved to be homogeneous in the Tiselius apparatus and in the ultracentrifuge. Its molecular weight was estimated to be 2.3 million. It was found that 1 mole of fatty acid synthetase at 25°C and pH 6.8 incorporates 3200 moles of malonyl-CoA into fatty acids per minute.

The synthesis of fatty acids from malonyl-CoA requires TPNH as a reducing agent and small amounts of acetyl-CoA in analogy with the avian and mammalian enzymes first studied by Wakil and Ganguly<sup>43</sup> and by Brady<sup>44</sup>. The yeast enzyme synthesizes a mixture of palmityl- and stearyl-CoA, according to equation (3):



(where  $n = 7$  or 8)

Acetyl-CoA serves as "primer" of the process. Its C<sub>2</sub>-unit is recovered only in the methyl end of the fatty acid produced, indicating that C<sub>2</sub>-units from malonyl-CoA are added to the acetyl residue during the synthetic reaction. In its function as "primer", acetyl-CoA can be replaced by homologous saturated acyl-CoA compounds but not by their oxidation products, identified in studies on fatty acid oxidation. Furthermore, all attempts to find low molecular weight intermediates of the synthesis were unsuccessful.

The explanation for these puzzling observations was our discovery that the transformation of malonyl-CoA into fatty acids is achieved through intermediates which are covalently bound to sulphhydryl groups of the synthetase<sup>42</sup>. We found that two different types of sulphhydryl groups have carrier function in the synthetic process. We denoted them as "central" and "peripheral" sulphhydryl groups for purposes of differentiation. In the following scheme of fatty acid synthesis (*Figure 18*), first presented in 1961<sup>42,45</sup> they are distinguished by bold faced and normal print.

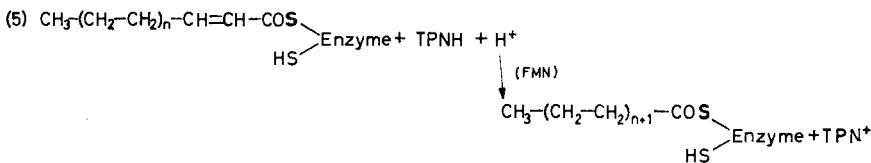
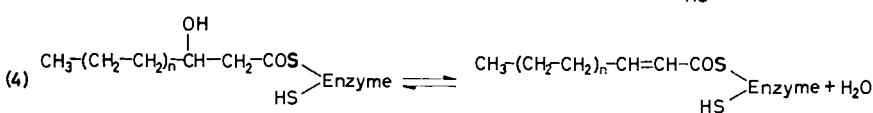
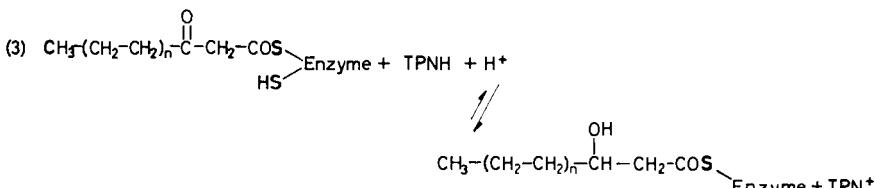
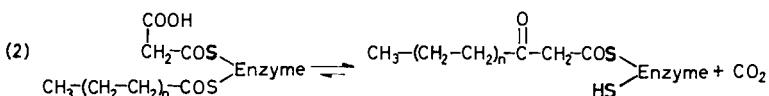
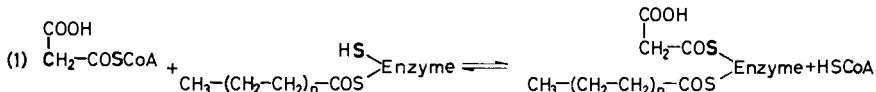
The synthetic process is initiated by the transfer of an acetyl residue from acetyl-CoA to the "peripheral" sulphhydryl group, designated as "priming reaction". It is followed by the transfer of a malonyl residue from malonyl-CoA to the "central" sulphhydryl group. The next step is a condensation between the enzyme-bound acetyl and malonyl groups resulting in the formation of acetoacetyl-enzyme with the concomitant liberation of carbon dioxide. The stepwise conversion of the  $\beta$ -keto acid into the saturated acid is accomplished by way of its reduction by TPNH to D(-)- $\beta$ -hydroxybutyryl-enzyme, followed by dehydration to crotonyl-enzyme and another TPNH-linked reduction to form the saturated butyryl-enzyme. In the second reduction step flavine mononucleotide (FMN) serves as hydrogen carrier. All the acyl residues involved in these chemical transformations are bound to the "central" sulphhydryl group. At the stage of the saturated acid finally the butyryl group is transferred to the "peripheral" sulphhydryl group, thus liberating the "central" sulphhydryl group for introduction of the next malonyl residue. The reaction cycle can then proceed again, starting with

## F. LYNEN

### Priming reaction:



### Chain lengthening reactions:



### Terminal reaction:

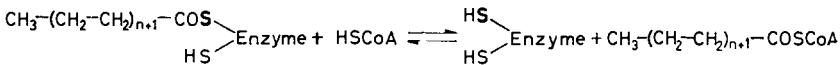


Figure 18. The mechanism of fatty acid synthesis

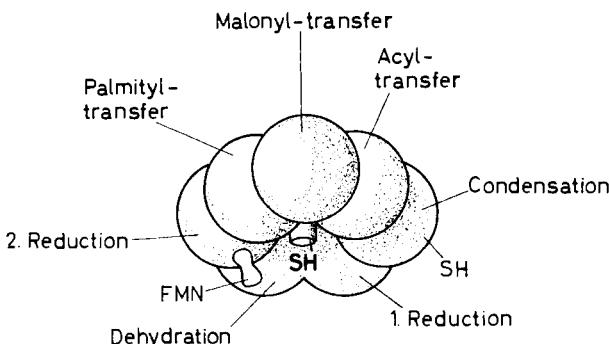
butyryl malonyl-enzyme, and is repeated until long chain saturated fatty acids with 16 or 18 carbon atoms are formed. In the terminal reaction step the acyl residue of palmitoyl- or stearoyl-enzyme is transferred from the "central" sulphhydryl group to coenzyme A with the formation of palmitoyl- or stearoyl-CoA and the regenerated enzyme. The free enzyme can again react with acetyl- and malonyl-CoA thereby reinitiating the entire process.

This whole sequence of reactions is accomplished by a multienzyme complex. Its functional unit was proposed to consist of a combination of seven different enzymes arranged around the "central" sulphhydryl group in such a manner, that the intermediates bound covalently to this group can

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

come into close contact with the active sites of the participating enzymes (*Figure 19*).

We have now obtained experimental evidence that the "central" SH-group belongs to an individual structural element<sup>46</sup>. As we originally assumed in our model, it is not part of one of the enzymic components. This means that the architecture of the fatty acid synthetase from yeast resembles the structure of the analogous enzyme system from bacteria. As the elegant experiments of Vagelos and his associates<sup>47</sup> with the *E. coli* system have demonstrated, the "central" sulphhydryl group, which carries the fatty acid intermediates, is bound to a readily dissociable protein of molecular weight about 9500, designated "acyl carrier protein". The bacterial enzyme system is not arranged in a stable multienzyme complex. Using standard methods of protein fractionation, it was possible to separate this enzyme system into enzymatically active individual components. A similar "acyl carrier protein" has been reported to be present in the corresponding enzyme system from plants<sup>48</sup>.



*Figure 19.* Hypothetical structure of the multienzyme complex of fatty acid synthetase. The seven enzyme units shown refer to the seven reactions (see *Figure 18*) occurring in fatty acid synthesis

Our repeated trials to split the multienzyme complex of yeast into its sub-units with retention of the individual enzyme activities were without success. In order to split the complex, it was necessary to use 0.2 M sodium desoxycholate or 6 M urea which are such drastic conditions that most of the individual enzyme activities of fatty acid synthesis were lost<sup>49</sup>. In its stability the multienzyme complex of yeast fatty acid synthetase resembles the avian and mammalian synthetase complexes<sup>50,51</sup>.

From the standpoint of chemical mechanism all fatty acid synthetases studied so far seem to catalyze the same reaction sequence. One minor dissimilarity concerns the terminal reaction, which yields palmitoyl- and stearoyl-CoA with the yeast synthetase but free palmitate with the animal and bacterial enzyme systems. This difference may be due to the intervention of a hydrolytic enzyme which replaces the acyl transferase in the terminal reaction.

The structural organization of the fatty synthetase from yeast was verified by electron microscopy (*Figure 20*). By use of the negative staining technique

F. LYNEN

with phosphotungstic acid Hofschneider was able to recognize single particles of oval shape surrounded by an equatorial ring. The longitudinal diameter of the particles is 250 Å, their cross diameter 210 Å. Unfortunately we cannot yet translate the information from electron micrographs into known schemes of structure. However, a structure composed of three circular sub-units fitting together seems possible. From the results of the chemical studies we have good evidence which supports the concept that each particle of molecular weight 2.3 million is composed of three functional assemblies.

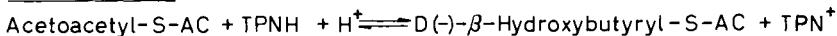
Malonyl-transfer:



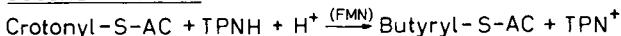
Palmityl-transfer:



First reduction:



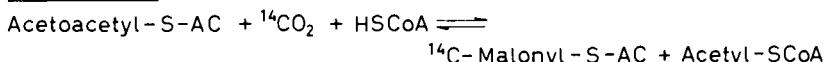
Second reduction:



Dehydration:



Condensation:

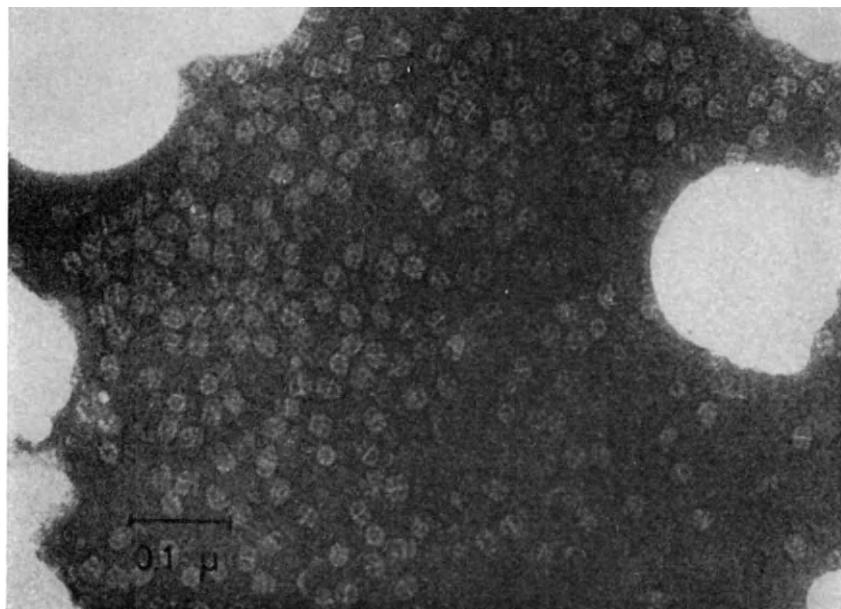


Where  $\text{S-AC} = \text{S-CH}_2\text{-CH}_2\text{-NH-CO-CH}_3$

Figure 21. Summary of assay reactions with model substrates for the determination of the enzymatic components of fatty acid synthetase

In order to demonstrate the manifold catalytic activities attributed to the synthetase, we used model substrates in which the carboxylic acid intermediates of fatty acid synthesis were bound to pantetheine or *N*-acetyl-cysteamine (Figure 21). Lacking the strong covalent bond to the "central" sulphhydryl group of the natural substrates, the affinity of these model substrates for the component enzymes is rather small. The defect can be circumvented, however, by employing high concentrations of the model substrates<sup>42</sup>. Studying the bacterial enzyme system, the natural substrates, that is, the carboxylic acid intermediates bound to the "acyl carrier protein" could be used<sup>52-55</sup>.

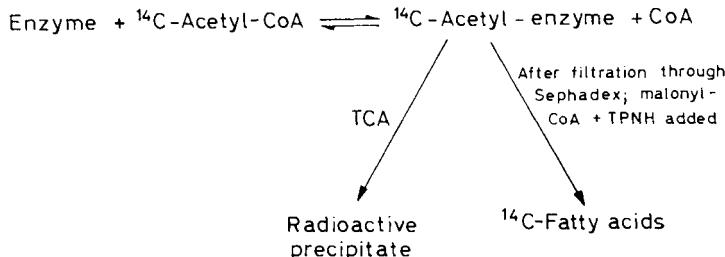
In addition to the experiments with model substrates we also used stoichiometric amounts of the yeast enzyme and demonstrated the enzyme bound intermediates directly<sup>45</sup>. As an example, short incubation of labelled 1-<sup>14</sup>C-acetyl-CoA with synthetase led to the formation of the radioactive acetyl-enzyme (Figure 22). It could be precipitated with trichloroacetic



*Figure 20.* Electron micrograph of the purified fatty acid synthetase from yeast

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

acid (TCA) with retention of radioactivity, indicating that the labelled acetyl group was covalently linked to the protein. When the radioactive acetyl-enzyme was separated from excess substrate by passing the reaction mixture



*Figure 22.* Scheme of the formation and transformation of acetyl-enzyme

over Sephadex and then incubated with malonyl-CoA and TPNH, more than 90 per cent of the protein bound acetic acid could be recovered in the fatty acids formed.

If the incubation mixture of labelled  $1-^{14}\text{C}$ -acetyl-CoA and synthetase was supplemented with malonyl-CoA a radioactive acetoacetyl-enzyme was formed. This could also be precipitated with trichloroacetic acid and yielded  $3-^{14}\text{C}$ -acetoacetate after mild alkaline hydrolysis. Our trials to isolate the native acetoacetyl-enzyme, employing the Sephadex technique failed, because acetoacetyl-enzyme is slowly decomposed spontaneously to free acetoacetate. In these experiments we measured the equilibrium constant of the formation of acetoacetyl-enzyme and found the following value<sup>12</sup>:

$$K_{\text{eq}} = \frac{[\text{acetoacetyl-enzyme}] \times [\text{CoA}]^2 \times [\text{CO}_2]}{[\text{acetyl-CoA}] \times [\text{malonyl-CoA}] \times [\text{enzyme}] \times [\text{H}^+]} = 2 \times 10^5 \text{ (0°C)}$$

Eliminating the  $\text{H}^+$ -concentration in this equation the equilibrium constant at pH 7.0 has the value:

$$K'_{\text{eq}} = \frac{[\text{acetoacetyl-enzyme}] \times [\text{CoA}]^2 \times [\text{CO}_2]}{[\text{acetyl-CoA}] \times [\text{malonyl-CoA}] \times [\text{enzyme}]} = 2 \times 10^{-2}$$

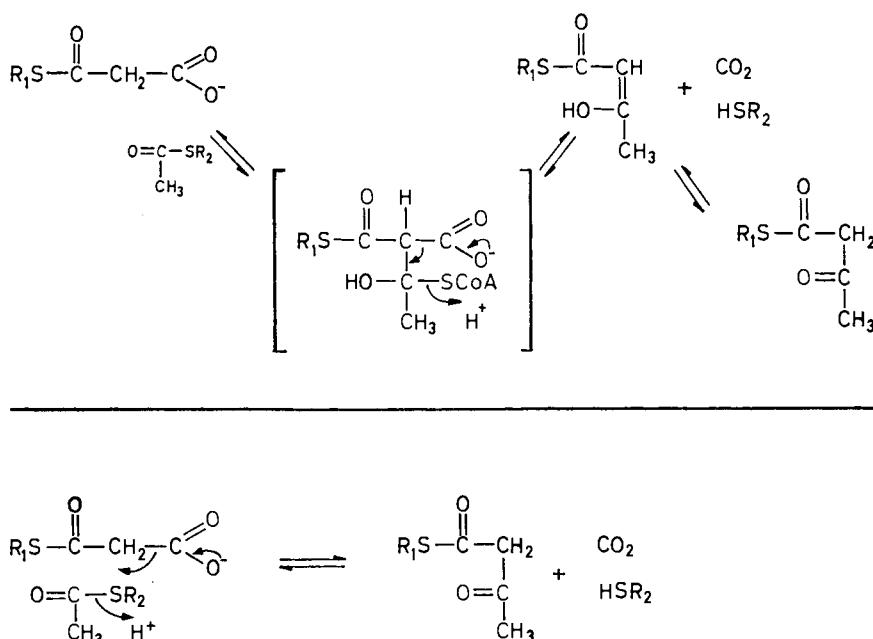
If we compare this equilibrium constant (condensation with malonyl-CoA) with the equilibrium constant of the thiolase reaction (condensation with acetyl-CoA, cf. page 144), which also generates an acetoacetyl thioester, the great thermodynamic advantage gained by the decarboxylation accompanying the condensation with malonyl-CoA becomes evident.

The favourable shift in the equilibrium position of the condensation reaction is ultimately due to the delivery of energy by ATP. At the expense of one pyrophosphate bond acetyl-CoA is first bound to  $\text{CO}_2$  as a carrier. After replacement of coenzyme A by the "central" sulphhydryl group of the enzyme complex, condensation occurs with the acetyl residue bound to the "peripheral" sulphur. The condensation is accompanied by the release of carbon

F. LYNEN

dioxide and the cleavage of one thioester bond. Because both reactants are in an "activated" state, the condensation reaction becomes very efficient.

From a chemical standpoint the condensation reaction may be classified as an acylation of malonic ester. The methylene group of the malonyl thioester, which is known to be more nucleophilic than the methyl group of an acetyl thioester, adds to the electrophilic carbonyl carbon of the sulphur bound carboxylic acid, as shown in the upper half of *Figure 23*. The intermediate formed is converted into the  $\beta$ -keto acyl derivative by the subsequent elimination of mercaptan and carbon dioxide. An alternative reaction mechanism might be the concerted process shown on the lower half of *Figure 23*. A decision between the two mechanisms may be reached by tracer experiments with heavy water.



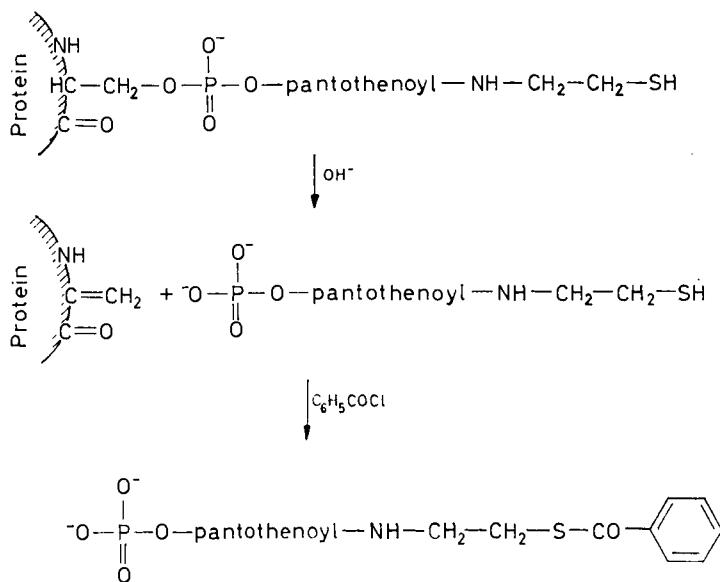
*Figure 23.* Mechanism of the formation of  $\beta$ -keto acids from malonyl thioester

Investigation of the chemical nature of the two types of sulphhydryl groups involved in fatty acid synthesis first led to the identification of the "peripheral" one as belonging to cysteine. In addition, we have obtained experimental evidence, that this sulphhydryl group is located in the condensing enzyme components of the multienzyme complex<sup>56</sup>.

The carrier of the "central" sulphhydryl group of the yeast fatty acid synthetase was recently identified as 4'-phosphopantetheine bound through phosphodiester linkage with the hydroxyl group of a serine residue of the polypeptide<sup>57</sup>. This kind of attachment was discovered by Vagelos<sup>58</sup> and by Wakil<sup>59</sup> in studies on the chemical structure of the acyl carrier protein of *E. coli*, and was also found in the mammalian fatty acid synthetase complex<sup>51</sup>.

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

Our evidence for the occurrence of 4'-phosphopantetheine stems from experiments in which the purified fatty acid synthetase of yeast was heated at pH 12 at 98° for one hour, following the procedure described by Vagelos<sup>58</sup>. Under these conditions the protein released a low molecular weight compound, which after benzoylation and further purification was identified as S-benzoyl-4'-phosphopantetheine by chemical analysis and by comparison with the authentic compound (*Figure 24*). The release of 4'-phosphopantetheine by mild alkaline treatment is in accord with the concept of an elimination reaction in the  $\beta$ -position of a polypeptide bound serine.



*Figure 24.* Schematic representation of the identification of protein bound 4'-phosphopantetheine

The presence of 4'-phosphopantetheine is further supported by experiments with yeast cells grown in a medium containing <sup>14</sup>C-labelled pantothenic acid. Wells, as a guest in my laboratory, isolated the fatty acid synthetase from these cells and found the purified enzyme to contain three moles of radioactive pantothenate covalently bound per mole of enzyme<sup>57</sup>.

The attachment through 4'-phosphopantetheine provides a flexible arm for the "central" sulphhydryl group, conceivably permitting rotation of the latter between the various enzymes. In this manner it is easily possible to bring the fatty acid intermediates, bound covalently to this sulphhydryl group, in close contact with the active site of each component enzyme which has only limited freedom of motion in the stable multienzyme complex. This is schematically illustrated in *Figure 25*, where the circles should indicate the active sites of the participating enzymes.

In the course of our investigations which led to the identification of the

F. LYNEN

"peripheral" and "central" sulphhydryl groups we found to our great surprise that acetate and malonate are not bound to the enzyme complex exclusively via sulphur atoms<sup>60</sup>. In these experiments we used the lability of thiol esters towards performic acid as a tool for the characterization.

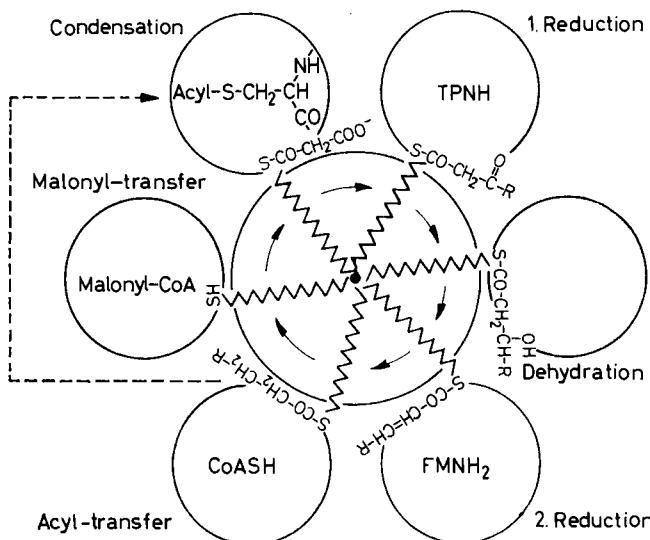
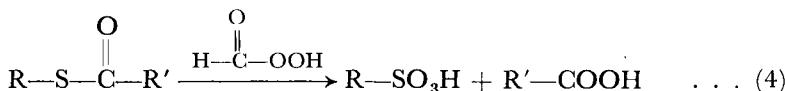


Figure 25. Reaction scheme illustrating individual events during fatty acid synthesis on the multienzyme complex

Performic acid oxidizes thioesters to the corresponding sulphonic acids with release of the carboxylic acids (equation 4),



In Table 2 the results of an experiment with radioactive  $^{14}\text{C}$ -acetyl enzyme are shown. The labelled enzyme was prepared by the interaction of constant amounts of enzyme with varying concentrations of  $1\text{-}^{14}\text{C}$ -acetyl-CoA, followed by precipitation with trichloroacetic acid and measuring the protein bound radioactivity. From this table it can be seen that the transfer of radioactive acetate to the enzyme depends on the concentration of acetyl-CoA used, as expected, if the acetyl transfer is a reversible process. However, in the whole concentration range studied, only about 50 per cent of the radioactive acetyl groups bound to the protein were released by treatment with performic acid.

From this and other experiments we came to the conclusion that the acyl transfer to the multienzyme complex is initiated with the transfer to some non-sulphhydryl acceptor group X in the protein. As illustrated in Figure 26 malonyl and acetyl residues are transferred from group X to the "central"

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

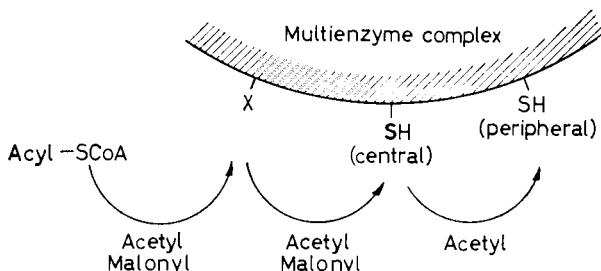
*Table 2.* Formation of  $^{14}\text{C}$ -acetyl enzyme and release of  $^{14}\text{C}$ -acetic acid by oxidation with performic acid

Each sample contained in 1.0 ml of 0.1 M K-phosphate, pH 6.5: 2 mg of yeast fatty acid synthetase (sp. activity 850 milliunits), 10  $\mu\text{moles}$  of cysteine and  $^{14}\text{C}$ -acetyl ( $24.8 \times 10^6$  c.p.m./ $\mu\text{mole}$ ) in the concentrations listed in the table. The mixture was incubated for 5 min at 0°C and the reaction stopped by the addition of 0.3 ml of 3 M trichloroacetic acid. After careful washing the precipitated protein was dissolved in 0.5 ml of 98 per cent formic acid, 0.2 ml aliquots dried on strips of Whatman 1 paper and the radioactivity measured before and after 12 h exposure to an atmosphere of performic acid. Before measurement in the scintillation counter the paper strips were dried in vacuum over potassium hydroxide. The radioactivity values listed in the table are based on 2 mg protein.

$^{14}\text{C}$ -Acetyl-CoA ( $\times 10^{-6}$ M)	Protein bound radioactivity (c.p.m.)		Stable radioactivity (%)
	Direct	After oxidation	
2	15 623	7 825	50
4	19 582	9 100	46.5
6	23 847	10 045	42
10	29 325	13 260	45
15	38 590	17 270	45
20	39 830	17 731	45
30	40 817	18 921	46
40	45 437	20 741	45.5
50	50 113	22 921	45.5

sulphydryl group. The further transfer to the "peripheral" sulphydryl group is specific for acetate and its higher homologues.

According to our scheme, the acyl residue of malonyl enzyme is bound to both group X and the "central" sulphydryl group. In agreement with this

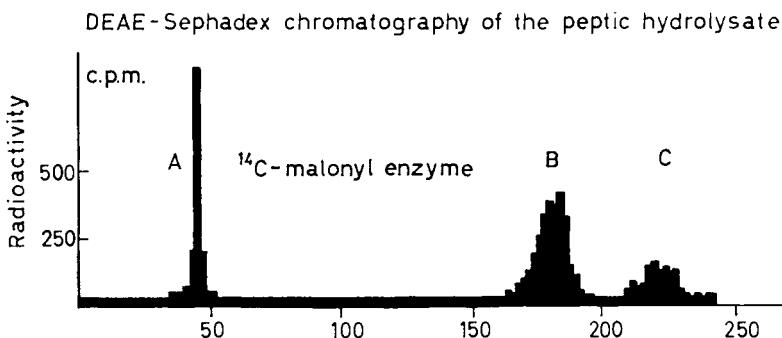


*Figure 26.* Specificity of acyl transfer to the different acceptor groups on the multienzyme complex

assumption, more than one radioactive acyl peptides were isolated from the peptic hydrolysate of  $3-^{14}\text{C}$ -malonyl enzyme by chromatography on DEAE-Sephadex. This is illustrated in *Figure 27*. The fastest moving fraction, signified A, was split by performic acid, whereas fractions B and C were stable to the same treatment. From the acid hydrolysate of malonyl peptide A, cysteamine and  $\beta$ -alanine were identified by ion exchange chromatography<sup>60</sup>. Since cysteine was absent, this peptide is derived from the polypeptide area around the "central" sulphydryl group.

F. LYNEN

The radioactivity of the fractions B and C was due to the presence of a  $^{14}\text{C}$ -malonyl heptapeptide and a  $^{14}\text{C}$ -malonyl pentapeptide which could be purified by chromatographic procedures. The acyl pentapeptide contained serine besides histidine, glycine, alanine, and leucine (C-terminal position).



*Figure 27.* Chromatography of the peptic hydrolysate of  $^{14}\text{C}$ -malonyl-enzyme on DEAE-Sephadex. For the preparation of  $^{14}\text{C}$ -malonyl enzyme 1.6 g of synthetase (sp. activity 550 milliunits/mg) and 28  $\mu\text{moles}$  of 3- $^{14}\text{C}$ -malonyl-CoA ( $1.2 \times 10^6$  c.p.m./ $\mu\text{mole}$ ) in 120 ml of 0.05 M K-phosphate buffer, pH 6.5, were incubated for 2 min at 22°C. The reaction was stopped by the addition of 6 ml of 3 M trichloroacetic acid. After careful washing, the precipitated protein was suspended in 2 l of 0.01 N HCl and digested by incubation with 160 mg of crystalline pepsin at room temperature. After evaporation of the water in vacuum the residue was dissolved in 7 ml of pyridine acetate buffer, pH 6.5. Insoluble material was removed by centrifugation, and the clear solution chromatographed on a column of DEAE-Sephadex (2.1  $\times$  150 cm). The eluting solvent until fraction 70 (volume of fraction: 6 ml) was 0.05 M pyridine acetate buffer, pH 6.25; thereafter a gradient was employed. For this purpose 0.2 M acetic acid was continuously introduced into a mixing chamber containing 200 ml of 0.05 M pyridine acetate buffer, pH 6.25

The acyl heptapeptide was similar but contained, in addition, glycine and glutamic acid. These results seem to indicate that group X is identical with the hydroxyl group of serine. Further experiments should indicate whether the carrier group X is connected with the enzyme components that catalyze acyl transfer reactions.

Another unsolved problem is why fatty acid synthetase produces mainly palmitic and stearic acid but no carboxylic acids of shorter chain length. Earlier we assumed that this might be due to the specificity of the enzymic component which terminates the process by transferring the fatty acid radical from the enzyme to coenzyme A<sup>12</sup>. However, when we studied the relationship between rate of acyl transfer and chain length of the acid, it was found to our great surprise that the saturated acids from  $\text{C}_6$  to  $\text{C}_{20}$  were transferred at nearly equal rates<sup>61</sup>. Consequently other factors must be responsible for stopping the synthesis at the stage of the  $\text{C}_{16}$  and  $\text{C}_{18}$  acids.

#### *Physiological regulation of fatty acid synthesis*

In conclusion I would like to discuss briefly our present views about the rate control of fatty acid synthesis. This will bring us back to the biotin

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

enzyme acetyl-CoA carboxylase, the step recognized to be rate-limiting in fatty acid synthesis in warmblooded organisms<sup>62</sup>. As has been shown in several laboratories, one of the unique features of acetyl-CoA carboxylase from animal sources is its activation by tri- and dicarboxylic acids, especially citrate. Extensive studies of this stimulatory effect have revealed that the increased activity is connected with an aggregation of the enzyme (cf. ref. 62). Another interesting property of the acetyl-CoA carboxylase is its inhibition by long-chain acyl-CoA derivatives, as was discovered by Bortz. It was found that the inhibition increased with growing chain-length of the fatty acid radical. Numa carried out systematic kinetic studies on the inhibition by palmitoyl-CoA with the purified enzyme. He found that the inhibition was competitive with regard to citrate, the enzyme activator<sup>62,63</sup>.

Our studies have been restricted to the properties of the enzyme, *in vitro*, and must be cautiously interpreted with regard to the physiological regulation of fatty acid synthesis. Nevertheless, it is conceivable that our findings might represent part of the cellular control mechanism. Raised levels of fatty acids in the blood are associated with starvation and diabetes, conditions in which fatty acid synthesis is known to be almost fully blocked. Further, it could be shown in normal animals that a diet rich in fat or the infusion of chylomicrons led to a drastic inhibition of fatty acid synthesis. Related to these observations it was found that the level of the higher acyl-CoA compounds in liver is markedly increased under all conditions of depressed fatty acid synthesis (*Table 3*). Therefore it is tempting to speculate

*Table 3.* Long chain acyl-CoA content of rat liver

<i>Status of rats</i>	<i>Acyl-CoA per g fresh liver (mμ moles)</i>	<i>Researchers</i>
Normal	14.9 ± 0.5	{ Bortz and
Starved for 24 h	57.7 ± 7.9	Lynen (1963)
Normal	52.8 ± 8.6	{ Tubbs and
Starved for 48 h	110.0 ± 28	Garland
After fat feeding	135.0 ± 23	(1963, 1964)
Diabetic	92.0 ± 20	
Normal	14.5 ± 2.6	{ Wieland;
Diabetic	58.2 ± 10.6	Felts <i>et al.</i>
Diabetic and insulin-injected	16.0 ± 2.3	(1964)

that fatty acid synthesis is under typical feedback control. As can be seen from the pattern of lipogenesis (*Figure 28*), the inhibitory long-chain acyl-CoA derivatives represent the last molecules in the synthetic sequence before subsequent incorporation into the "complex lipids". Furthermore, the inhibition would affect the enzyme acetyl-CoA carboxylase, the point at

which fatty acid synthesis branches off from the many other reaction paths of acetyl-CoA.

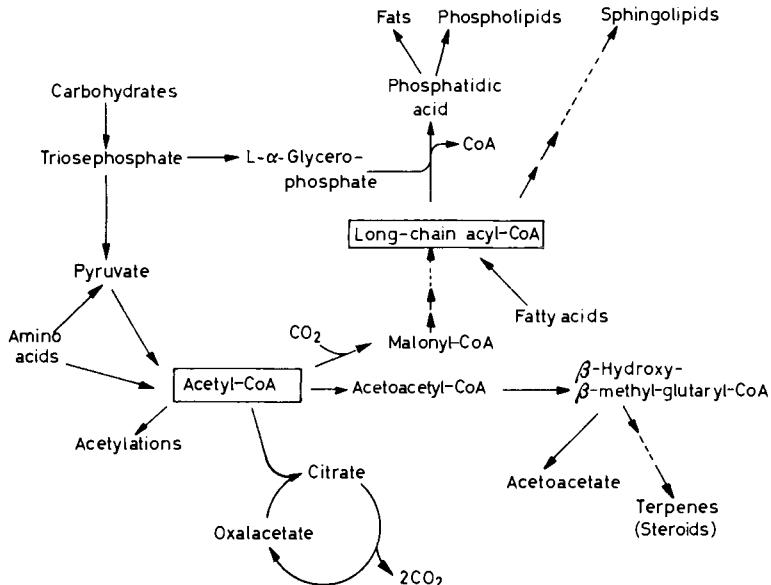


Figure 28. The biological pathway of lipogenesis

### Biosynthesis of more complex polyacetate structures

Let us finally examine the implications of the studies on fatty acid synthesis for the biosynthesis of the more complex polyacetate structures. It has often been remarked that no intermediates in the postulated biogenesis have been found<sup>3</sup>. Therefore one might imagine that these substances are also synthesized on multienzyme complexes with covalently bound intermediates<sup>40,42</sup>. With this view one could reconcile to the concept of Ehrensvärd<sup>64</sup> and of Woodward<sup>39</sup> which they developed for the biosynthesis of polycyclic quinones and macrolides. Such enzyme complexes represent, so to speak, the matrices on which the raw materials are assembled piece by piece. The product would be released from the enzyme complex only in the final form, thus explaining the inability to detect intermediates. This supposition is illustrated in Figure 29 for the case of 6-methylsalicylic acid synthesis. Looking at the depicted reaction sequence, we see acyl transfer reactions, condensations, reductions, and dehydrations as in the case of fatty acid synthesis. What is responsible for the fact that the same kind of enzyme activities can give rise to fatty acids on the one hand and 6-methylsalicylic acid on the other? Is this solely the consequence of the specificities of the component enzymes, or does the specific architecture of the multienzyme complex, by determining the order of reactions, play a decisive role? This is

BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

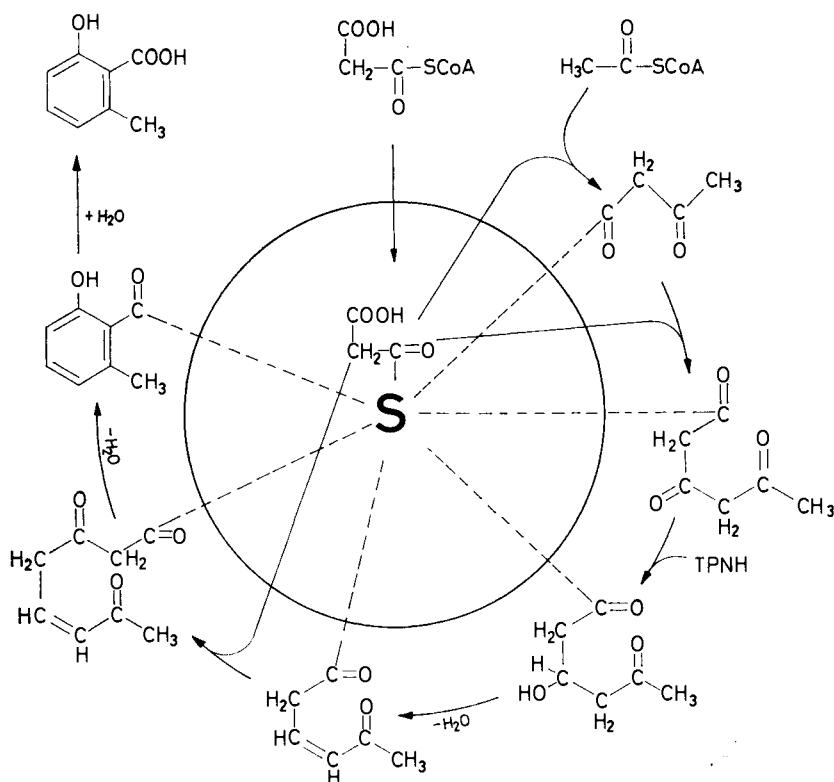


Figure 29. Hypothetical reaction scheme of 6-methylsalicylic acid synthesis on a multienzyme complex

an important question which research on the enzyme level will have to answer.

Finally, it is a pleasant duty to mention the enthusiasm and devotion of all my associates who have made our work possible, and whose names are given in the various references. I also thank Professor W. W. Wells for his kind help in the preparation of this manuscript.

## References

- <sup>1</sup> L. Ruzicka. *Experientia* **9**, 357 (1953).
- <sup>2</sup> A. J. Birch. In *Fortschritte der Chemie organischer Naturstoffe* (Ed. L. Zechmeister), Vol. 14, p. 186, Springer-Verlag, Wien, 1957.
- <sup>3</sup> J. H. Richards, and J. B. Hendrickson. *The Biosynthesis of Steroids, Terpenes, and Acetogenins*, W. A. Benjamin, Inc., New York-Amsterdam, 1964.
- <sup>4</sup> S. Chaykin, J. Law, A. H. Phillips, T. T. Tchen, and K. Bloch. *Proc. Natl. Acad. Sci. U.S.* **44**, 998 (1958).
- <sup>5</sup> F. Lynen, H. Eggerer, U. Henning, and I. Kessel. *Angew. Chem.* **70**, 738 (1958).
- <sup>6</sup> B. L. Archer, D. Barnard, E. G. Cockbain, J. W. Cornforth, R. H. Cornforth, and G. Popják. *Proc. Roy. Soc. Ser. B.* **163**, 519 (1965/66).

F. LYNEN

- <sup>7</sup> H. Eggerer. *Chem. Ber.* **94**, 174 (1961).
- <sup>8</sup> W. S. Johnson, and R. A. Bell. *Tetrahedron Letters* **12**, 27 (1960).
- <sup>9</sup> F. Lynen, B. W. Agranoff, H. Eggerer, U. Henning, and E. M. Mösllein. *Angew. Chem.* **71**, 657 (1959).
- <sup>10</sup> U. Henning, E. M. Mösllein, B. Arreguin, and F. Lynen. *Biochem. Z.* **333**, 534 (1961).
- <sup>11</sup> B. L. Archer, B. G. Audley, E. G. Cockbain, and G. P. MacSweeney. *Biochem. J.* **89**, 565 (1963).
- <sup>12</sup> F. Lynen. *Angew. Chem.* **77**, 929 (1965).
- <sup>13</sup> A. Eschenmoser, L. Ruzicka, O. Jeger, and D. Arigoni. *Helv. Chim. Acta* **38**, 1890 (1955).
- <sup>14</sup> P. Karrer, and A. Helfenstein. *Helv. Chim. Acta* **14**, 78 (1931).
- <sup>15</sup> J. W. Cornforth, R. H. Cornforth, C. Donninger, and G. Popják. *Proc. Roy. Soc. Ser. B.* **163**, 492 (1965/66).
- <sup>16</sup> A. F. Wagner, and K. Folkers. *Adv. Enzym.* **23**, 471 (1961).
- <sup>17</sup> I. F. Durr, and H. Rudney. *J. Biol. Chem.* **235**, 2572 (1960).
- <sup>18</sup> J. Knappe, E. Ringelmann, and F. Lynen. *Biochem. Z.* **332**, 195 (1959).
- <sup>19</sup> J. Rétey, and F. Lynen. Unpublished experiments.
- <sup>20</sup> M. Lindberg, C. Yuan, A. de Waard, and K. Bloch. *Biochemistry* **1**, 182 (1962).
- <sup>21</sup> L. Jaenicke, and F. Lynen. In *The Enzymes* (Eds. P. D. Boyer, H. Lardy, and K. Myrbäck) Vol. III, p. 3, Academic Press, New York (1960).
- <sup>22</sup> F. Lynen. *Fed. Proc.* **12**, 683 (1953).
- <sup>23</sup> U. Gehring. Thesis, University of Munich (1964).
- <sup>24</sup> U. Gehring, and F. Lynen. Unpublished results.
- <sup>25</sup> H. Beinert, and P. G. Stansly. *J. Biol. Chem.* **204**, 67 (1953).
- <sup>26</sup> J. D. Brodie, G. Wasson, and J. W. Porter. *J. Biol. Chem.* **238**, 1294 (1963).
- <sup>27</sup> D. Bloomfield, and K. Bloch. *J. Biol. Chem.* **235**, 337 (1960).
- <sup>28</sup> N. L. R. Bucher, K. McGarrahan, E. Gould, and A. V. Loud. *J. Biol. Chem.* **234**, 262 (1959).
- <sup>29</sup> G. M. Tomkins, H. Sheppard, and I. L. Chaikoff. *J. Biol. Chem.* **201**, 137 (1953).
- <sup>30</sup> R. G. Langdon, and K. Bloch. *J. Biol. Chem.* **202**, 77 (1953).
- <sup>31</sup> I. D. Frantz, Jr., H. S. Schneider, and B. T. Hinkelmann. *J. Biol. Chem.* **206**, 465 (1954).
- <sup>32</sup> G. M. Tomkins, and I. L. Chaikoff. *J. Biol. Chem.* **196**, 569 (1952).
- <sup>33</sup> T. T. Hutchens, J. T. Van Bruggen, R. M. Cockburn, and E. S. West. *J. Biol. Chem.* **208**, 115 (1954).
- <sup>34</sup> R. G. Gould, V. L. Bell, and E. M. Lilly. *Radiation Res.* **5**, 609 (1956).
- <sup>35</sup> I. D. Frantz, Jr., and B. T. Hinkelmann. *J. Exptl. Med.* **101**, 225 (1955).
- <sup>36</sup> N. L. R. Bucher, P. Overath, and F. Lynen. *Biochem. Biophys. Acta* **40**, 491 (1960).
- <sup>37</sup> D. Regen, C. Riepertinger, B. Hamprecht, and F. Lynen. *Biochem. Z.* **346**, 78 (1966).
- <sup>38</sup> J. N. Collie. *J. Chem. Soc.* 122, 329 (1893); 787, 1806 (1907).
- <sup>39</sup> R. B. Woodward. *Angew. Chem.* **69**, 50 (1957).
- <sup>40</sup> F. Lynen, and M. Tada. *Angew. Chem.* **73**, 513 (1961).
- <sup>41</sup> F. Lynen. *J. Cell. Comp. Physiol.* **54**, suppl. 1, 33 (1959).
- <sup>42</sup> F. Lynen. *Fed. Proc.* **20**, 941 (1961).
- <sup>43</sup> S. J. Wakil, and J. Ganguly. *J. Am. Chem. Soc.* **81**, 2597 (1959).
- <sup>44</sup> R. O. Brady. *Proc. Natl. Acad. Sci. U.S.* **44**, 993 (1958).
- <sup>45</sup> F. Lynen. In Proceedings of the Robert A. Welch Foundation Conferences on Chemical Research, V. Molecular structure and biochemical reactions; p. 293, Houston, Texas (1962).
- <sup>46</sup> K. Willecke. Unpublished experiments.
- <sup>47</sup> P. W. Majerus, A. W. Alberts, and P. R. Vagelos. *Proc. Natl. Acad. Sci. U.S.* **53**, 410 (1965).
- <sup>48</sup> P. Overath, and P. K. Stumpf. *J. Biol. Chem.* **239**, 4103 (1964).
- <sup>49</sup> A. Hagen. Thesis, University of Munich (1963).
- <sup>50</sup> R. Bressler, and S. J. Wakil. *J. Biol. Chem.* **236**, 1643 (1961).
- <sup>51</sup> A. R. Larabee, E. G. McDaniel, H. A. Bakerman, and P. R. Vagelos. *Proc. Natl. Acad. Sci. U.S.* **54**, 267 (1965).
- <sup>52</sup> P. W. Majerus, A. W. Alberts, and P. R. Vagelos. *Proc. Natl. Acad. Sci. U.S.* **51**, 1231 (1964).
- <sup>53</sup> A. W. Alberts, P. W. Majerus, B. Talamo, and P. R. Vagelos. *Biochemistry* **3**, 1563 (1964).
- <sup>54</sup> S. J. Wakil, E. L. Pugh, and F. Sauer. *Proc. Natl. Acad. Sci. U.S.* **52**, 106 (1964).
- <sup>55</sup> F. Sauer, E. L. Pugh, S. J. Wakil, R. Delaney, and R. L. Hill. *Proc. Natl. Acad. Sci. U.S.* **52**, 1360 (1964).

## BIOSYNTHETIC PATHWAYS FROM ACETATE TO NATURAL PRODUCTS

- <sup>66</sup> F. Lynen. In *New Perspectives in Biology* (Ed: M. Sela), (B.B.A. Library Volume 4), p. 132, Elsevier Publishing Company, Amsterdam (1964).
- <sup>67</sup> W. W. Wells, J. Schultz, and F. Lynen. *Proc. Natl. Acad. Sci. U.S.* **56**, 633 (1966).
- <sup>68</sup> P. W. Majerus, A. W. Alberts, and P. R. Vagelos. *J. Biol. Chem.* **240**, 4723 (1965).
- <sup>69</sup> E. L. Pugh, and S. J. Wakil. *J. Biol. Chem.* **240**, 4727 (1965).
- <sup>70</sup> F. Lynen. In *Aspects of Yeast Metabolism*, Guiness Symposium, Dublin. To be published in 1967.
- <sup>71</sup> E. Schweizer. Thesis, University of Munich (1963).
- <sup>72</sup> S. Numa, W. M. Bortz, and F. Lynen. In *Advances in Enzyme Regulation* (Ed: G. Weber), Vol. III, p. 407, Pergamon Press, London (1965).
- <sup>73</sup> S. Numa, E. Ringelmann, and F. Lynen. *Biochem. Z.* **343**, 243 (1965).
- <sup>74</sup> G. Ehrensvärd, and S. Gatenbeck. In: *Hauptvorträge des XVII. Internationalen Kongress für Reine und Angewandte Chemie*, Vol. II, p. 73, Verlag Chemic, Weinheim/Bergstraße (1960).

## **SECTION 4: CHEMICAL TAXONOMY**

### **INTRODUCTORY REMARKS BY THE HONORARY PRESIDENT OF THE SECTION**

N. A. SØRENSEN

First of all I should like to express my warmest gratitude to The Organizing Committee for providing me an opportunity to have the very great honour to address you at this section on chemotaxonomy.

It is always a pleasure to address an enthusiastic audience. Regardless of opinion about chemotaxonomy, everyone must admit that this young discipline is crowded with enthusiastic scientists. In my opinion the term chemotaxonomy was selected somewhat arbitrarily. Taxonomy in the proper sense of the word is a science with one main goal "to achieve order in contemporary living plants". As a consequence of some of its methods, it has brought forward results also of phylogenetic importance. This was not aimed at, and so taxonomy as such dismiss all criticism not to be entirely phylogenetic.

Taxonomy is a remarkable science. Its techniques are, relative to other natural science approaches, relatively simple and rapid. As the only scientists who take care of all taxons it is understandable that taxonomists look sceptically down on all other sciences, which try to reach conclusions on a more or less restricted amount of material whether these be embryology, palynology, or chemotaxonomy. Of these related sciences, chemotaxonomy without doubt is in the worst position.

Professor Hegnauer in one of his reviews has requested that *the structure and distribution of all plant constituents must be investigated*. If we should have to follow Hegnauer on this point, we obviously have to admit that the complete chemical investigation of the first species is still not finished.

What characterizes our present situation is that each chemistry school is looking for one class of compounds, or one type of metabolite; all other compounds go down the sink. I suppose this looks to a real taxonomist as if plant classification was done by one botany school on leaves; by another on flowers, etc.

Then, is there any valid excuse for chemotaxonomy to exist? Personally I can think of at least three.

The classification of any plant should always be done on as many characters as possible. The characterization of the substances occurring in a

N. A. SØRENSEN

plant is thus a desired extension of the classical macroscopic and microscopic characters down to the level of molecules. In lichen systematics this extension has long been a necessity.

The famous botanist H. Hallier named *the elucidation of plant constituents and their distribution "Descriptive chemotaxonomy"*. As parallel developments in related genera in numerous cases lead to overlap in gross morphology, this "Descriptive chemotaxonomy" may be useful also for classification purposes if some chemical characters are very conservative and remain unchanged when morphology runs into overlapping. Cases of this type are known with different classes of chemical compounds and so chemists might be of some use.

I have to remind you, however, that in many cases the chemical characters are not at all constant. It should suffice to mention that the types of rape recently developed in Canada produce seeds devoid of erucic acid which was once one of the chemical markers of the Cruciferae. Anders Kiær has found mutant of *Capsella bursa pastoris* (L) Moench devoid of sulphur glycosides, another chemical marker of this family. Finally I may mention that *Conium maculatum* (L)—ill-fated since the days of Socrates—becomes devoid of all coniin when growing at higher altitudes, or, in other words, under long day conditions. Numerous phenomena connected with behaviour of plants under short or long day conditions are obviously due to a balance between synthesis and respiration. Edaphic influences, of which relative length of day and night is only one example, should not be allowed to blur our use of natural products in plant systematics.

It is very remarkable that Hallier in 1913 clearly defined another branch of chemotaxonomy the "Dynamic chemotaxonomy", comprising the biochemical pathways leading to parallel or diversified development of chemical characters. Hallier realized that this "Dynamic chemotaxonomy" could on one hand simplify the "Descriptive chemotaxonomy", on the other be a tool in phylogeny. And here I think we reach the second point where chemistry may be useful.

In recent years some botanists have introduced the "joke" talking about "equal weighting of characters". Although most of you are only chemists, you have long ago realized that a lot of the characters selected by botanists are absolutely decisive.

What I think is behind this "joke" is that just while the purpose of these characters is to divide for the sake of creating order, *these same characters* must be quite unsuitable to indicate any phylogenetic relations.

Ralph Alston of Texas—may I name him our "royal convertite" from classical botany to chemotaxonomy—divides natural products into: Basic Metabolites, Secondary Compounds, and Macromolecules. The remarkable achievements of our biochemical colleagues tell us not only that some of these macromolecules, *viz.* the nucleic acids and their functional translations the enzymic proteins decide whether "basic metabolites" shall accumulate as natural products—as do, *e.g.*, isocitric acid or sedoheptulose—as well as their transformations into our various types of secondary plant products. From this recent biochemical research emerges the evolutionary diversification of some fundamental proteins among them some enzymes. The aspect of "Dynamic chemotaxonomy" in the hands of these biochemists may develop into a new independent tool in phylogeny—and just the possibility

#### INTRODUCTORY REMARKS

of getting a new independent tool in this science will be highly appreciated—even if the evidence is fragmentary.

The third excuse I see is applied science, and as the A in IUPAC stands for Applied, I feel allowed to mention it.

Agricultural products present in crops to the extent of 10–30 per cent in pure form cost somewhere between 0·3–2 Sw. Kr./kg. This is a price range in which only products available through large scale petrochemistry may be manufactured. Many secondary compounds are useful to mankind either as such or as starting materials; practically none of them are available through any reasonable form of petrochemistry. The importance of the problem of concentration of secondary compounds may be enlightened by mentioning only two cases: Would there have been any steroid industry if some *Dioscorea* species had not happened to contain 4–5 per cent of diosgenin? Or where would prices or applications have found their limitations?

In this splendid review Dr. Karl Folkers mentioned the use of some alkaloids from *Vinca rosea* L. (*Catarranthus roseus* G. Don) against some types of cancer. Vincristin seems to be most promising, but the content in *Vinca rosea* is only 0·0025 per cent. Better principles may be found but at the moment the task of obtaining a plant source producing vincristine in reasonable yield is imperative.

In my opinion it is only if the natural product chemist develops the science of "Dynamic chemotaxonomy" that he will be fit to join the genetisist and the plant breeder for the development of a diversified natural products industry; I suppose this third excuse is a rather valid, although much ignored, excuse for chemotaxonomy.

# CHEMICAL CHARACTERS IN PLANT TAXONOMY: SOME POSSIBILITIES AND LIMITATIONS

R. HEGNAUER

*Laboratorium voor Experimentele Plantensystematiek, Leiden, Netherlands*

## INTRODUCTION

Chemical plant taxonomy or chemotaxonomy of plants may be defined as a scientific investigation of the potentialities of chemical characters for the study of problems of plant taxonomy and plant phylogeny. Plant taxonomy is the science of delimiting, describing and naming appropriately taxa† and arranging them in a natural system of plants.

Principles of chemotaxonomy were elaborated in the past century by A. P. De Candolle<sup>1</sup> and by Greshoff<sup>2</sup>. De Candolle put forward two postulates: (i) Plant taxonomy will be the most useful guide to man in his search for new industrial and medicinal plants; (ii) Chemical characteristics of plants will be most valuable to plant taxonomy in the future.

While the first postulate of De Candolle proved to be extremely fruitful and has been applied repeatedly when new sources of promising plant constituents are to be detected, his second postulate came to be accepted very slowly. Researchers like Rochleder<sup>3</sup>, Greshoff<sup>4</sup>, Rosenthaler<sup>5</sup>, Baker and Smith<sup>6</sup>, Wheldale<sup>7</sup>, Iwanow<sup>8</sup>, Colin<sup>9</sup>, Molisch<sup>10</sup>, McNair<sup>11</sup> and Weevers<sup>12</sup> were enthusiastic but rather isolated workers in the field of chemotaxonomy.

However, the fact that the first postulate of De Candolle was applied very successfully by generations of phytochemists forms an indirect proof of the validity of his second postulate. Some examples may serve to illustrate just the services plant taxonomy renders to chemists interested in distinct types of plant constituents. When the pharmaceutical industry became interested in plant steroids as starting materials for hormone synthesis, the search for suitable sources was essentially guided by taxonomic concepts. The genus *Strophanthus* was investigated first for cardenolides and its species proved, without exception, to accumulate members of this category of phytoconstituents. Thousands of species were screened for steroid saponins and in the taxa already known to contain them, *i.e.* in *Agavaceae*, *Dioscoreaceae* and *Liliaceae*, by far the highest frequency of occurrence was observed. In recent years the pregnane-derived alkaloids have begun to attract attention. Such alkaloid-like substances had been known for several years to be present in the apocynaceous genus *Holarrhena*. In this instance too, an alliance of genera included by taxonomists in the plant family *Apocynaceae* proved to be most promising for exploration. Very recently

† Taxon (plural: taxa) is the name for a taxonomic entity of unspecified rank; *i.e.* the term may be applied to any systematic entity (species, genus, family, etc.).

interest has developed in the steroid alkaloids of *Buxus sempervirens* L. Quite logically other species of *Buxus* and other genera of the small family of *Buxaceae* were explored for the same type of alkaloid-like substances. We are not surprised that such compounds were indeed detected in the genera *Pachysandra* and *Sarcococca* in spite of the fact that their species look quite different from *Buxus sempervirens*, the original source of this type of alkaloid-like substances. In my opinion, such a finding is a tribute to the work of generations of taxonomists endeavouring to elaborate a natural system of plants.

Of course, there are many not yet fully understood irregularities in the distribution of plant constituents. Steroidal sapogenins occur, e.g., in some genera of *Leguminosae*, *Solanaceae* and *Zygophyllaceae*, taxa which are distinctly not related to *Liliiflorae* and cardenolides are very erratically distributed over angiospermous plants. The famous French school of heteroside chemists founded by Bourquelot has detected many cases of sporadic occurrences of glycosides during the first four decades of this century. The intensive study of the causes of melanogenesis in fading plant tissues demonstrated, for instance, that this phenomenon may be caused in non-related taxa by one and the same constituent. Arbutin [*Ericaceae*, *Proteaceae*, species of *Pyrus* (*Rosaceae*), species of *Bergenia* (*Saxifragaceae*), *Lathyrus niger* (*Leguminosae*), some species of *Rubiaceae*] and aucubin [e.g. *Aucuba* (*Cornaceae*), *Eucommiaceae*, *Globulariaceae*, *Plantaginaceae*, *Scrophulariaceae*] are good examples of relatively wide-spread plant chromogenes. These facts led Bridel and Kramer<sup>13</sup> to deny any relationship between plant morphology and plant metabolism. After having isolated phlorizin from leaves of *Kalmia latifolia* (*Ericaceae*) they declared "il est intéressant de faire ressortir que le phlorizoside qui, jusqu'ici était regardé comme un principe spécifique de l'écorce de quelques Rosacées se rencontre également dans les feuilles et les fleurs de deux Ericacées, famille très éloignée des Rosacées au point de vue botanique. Cela prouve qu'ils n'existent guère des rapports entre les caractères botaniques et la composition chimique des plantes. D'autres exemples récents, notamment ceux qu'on tire de la répartition dans le règne végétal du monotropitoside, de l'aucuboside, de l'arbutoside et du picéoside, viennent renforcer cette opinion".

Contrary to this statement I expect chemical characters to be as valid in future for taxonomic work as are morphological ones. To reach this stage, however, our knowledge about plant metabolism and its resulting products still has to be considerably extended.

In traditional plant taxonomy the totality of morphological characters has always to be weighed and checked carefully when decisions with regard to delimitations and classification have to be made. For instance, Sympetalry is believed to be a most important character in dicotyledons but far less so in monocotyledons. The density and nature of the indument is a reliable taxonomic character in one group of plants but not at all in another. The same holds good for the structure of fruits and subterraneous organs and many other morphological characters. In taxonomy, generally, not one or a few characters but the total look of a taxon is most important. To this integral picture, without any doubt, metabolism contributes too. In my opinion, the answer given by Colin<sup>9</sup> as to which types of characters may be

## CHEMICAL CHARACTERS IN PLANT TAXONOMY

most useful for the study of problems of plant taxonomy can be accepted without restriction. He stated "Cela dépend sans doute du génie de la famille, comme disait Adanson, à qui l'on fait injure en s'appliquant à le disculper d'avoir méconnu le principe de la subordination des caractères . . ."

For most chemical characters, however, the overall information available at present does not yet suffice for a correct appreciation of their real contribution to the total look of a taxon. Therefore it is still impossible to judge appropriately their taxonomic importance. In many instances, of course, the taxonomic potentialities of chemical characters are seemingly apparent already.

### SOME TAXONOMIC POSSIBILITIES OF CHEMICAL CHARACTERS

#### **Chemical Characters as guides for classification**

The position of many taxa in the natural system of plants is still highly uncertain. This applies to all levels of taxonomic categories, e.g. species in a genus (example: *Matricaria inodora* L. in *Matricaria*, *Chrysanthemum* or *Tripleurospermum*), genera in a family (examples: *Morina* in *Dipsacaceae*; *Torriceillia* and *Corokia* in *Cornaceae*), families in an order (examples: *Hippuridaceae* in *Haloragales*; *Adoxaceae* in *Dipsacales*) and even orders in a class (example: *Taxales* in *Coniferales*). I chose three families of flowering plants to illustrate this point; the position attributed to them in 6 recent systems of angiosperms is given in *Table 1*.

*Table 1.* Position of three families in six recent systems of dicotyledons

Family	Placed in the following orders by:					
	Wettstein <sup>14</sup>	Pulle <sup>15</sup>	Cronquist <sup>16</sup>	Hutchinson <sup>17</sup>	Takhtajan <sup>18</sup>	Engler's <sup>19</sup> Syllabus
<i>Callitrichaceae</i>	Tricoccae	Callitrichales	Haloragales	Lythrales	Tubiflorae-Lamiales	Tubiflorae-Verbenineae
<i>Cornaceae</i>	Umbelliflorae	Apiales ( <i>=Umbelliflorae</i> )	Cornales*	Araliales†	Cornales§	Umbelliflorae
<i>Hippurideaceae</i>	Myrtales	Hippuridales	Haloragales	Lythrales‡	Myrtiflorae-Haloragales	Myrtiflorae-Hippuridineae

|| Derived from Solanales (*=Tubiflorae*).

\* Not allied to Umbelliflorae.

† Araliaceae but not Umbelliferae are included in this order.

‡ *Hippuris* is included in Haloragaceae by Hutchinson.

§ Preceding Araliales sensu Takhtajan (i.e. Araliaceae and Umbelliferae).

Varying interpretation and evaluation of morphological characters very often result in disagreement regarding classification. In such instances taxonomists as a rule look for characters other than morphological ones (see for instance Thorne<sup>20</sup>, Benson<sup>21</sup>, Davis and Heywood<sup>22</sup>). Generally anatomical (Solereider<sup>23</sup>, Metcalfe and Chalk<sup>24</sup>, Carlquist<sup>25</sup>), embryological (Maheshwari<sup>26</sup>), palynological (Wodehouse<sup>27</sup>, Erdtman<sup>28</sup>) and cytological (Darlington<sup>29</sup>, Manton<sup>30</sup>) characters are considered first. Sometimes they produce convincing evidence and sometimes they fail to do so.

In such situations chemical characters may become very useful guides to taxonomists. At present one important task of chemotaxonomy consists in procuring additional evidence in all cases of obscure relationships of plants.

I should like to illustrate this further for the three families already mentioned in *Table 1*. *Table 2* summarizes the principal present-day knowledge concerning their constituents.

*Table 2.* Some constituents of three families of plants of obscure relationship

Taxon	<i>Iridoid Heterosides*</i>	<i>Principal sugars (vegetative organs)*</i>	<i>Principal phenolic compounds</i>
<i>Callitrichaceae</i>	aucubin, catalpol	sucrose	flavones (probably); caffeic acid
<i>Cornaceae :</i> <i>Aucuba</i>	aucubin	sucrose	flavonols (probably); caffeic acid
<i>Cornus</i>	cornin	glucose, sucrose	flavonols; gallic and ellagic acids; leucoanthocyanins
<i>Corokia</i> <i>Grieselinia</i> <i>Mastixia</i>	cornin not present loganin, loganic acid	sucrose glucose, sucrose	leucoanthocyanins caffeic acid; flavonols caffeic acid
<i>Hippuridaceae</i>	aucubin, catalpol	stachyose	caffeic and ferulic acids; kaempferol and scopoletin (probably)

\*For most of the observations reported in these columns I am obliged to my collaborators Miss Fikenscher and Mr. Wiesfleier.

A glance at *Tables 1* and *2* demonstrates distinctly that in all instances chemical characters agree well with proposals already put forward by some taxonomists.

*Callitrichaceae* and *Hippuridaceae* fit chemically very well in the alliance of *Tubiflorae* (compare proposal of Pulle) and differ fundamentally from members of *Myrtiflorae* (including *Haloragales* and *Lythrales*).

*Cornaceae* represent perhaps a rather heterogeneous family<sup>31</sup>. As far as chemical information is available the latter indicates a rather intimate relationship with the saxifrageous stock and the sympetalous families of the orders *Contortae* and *Rubiales*. This makes rather acceptable an intermediate position between *Saxifragales* (or *Rosales*) and *Contortae-Rubiales* for *Cornaceae* and allied families and points distinctly against an association with *Araliaceae* and *Umbelliferae*<sup>31,32</sup>. Direct derivation of *Cornales* from *Rosales* was proposed by Cronquist<sup>16</sup>.

It is my viewpoint that in every instance, where fundamental disagreements regarding relationship and classification of taxa exist between experienced taxonomists, thorough phytochemical investigations may result in a better understanding and a re-evaluation of all available facts. In this respect I should like to draw your attention to the monotypic genus *Simmondsia* which is usually, but doubtfully, included in the already mentioned family of *Buxaceae*. If a chemist were to investigate the alkaloids present in *Simmondsia californica* Nutt. he would render a most valuable service to plant taxonomy.

### Chemical characters as aids in delimitations

Taxonomists endeavour to delimit taxa in such a manner that they really represent natural entities. In many instances, however, it is far from easy to

## CHEMICAL CHARACTERS IN PLANT TAXONOMY

conceive true naturalness, *i.e.* to grasp "le génie du taxon". To illustrate this point I would like to summarize two different concepts of liliaceous and amaryllidaceous plants. Traditionally *Liliaceae* are characterized by having hypogynous flowers with a showy perianth, 3 + 3 stamens and a pistillum composed of 3 carpels and *Amaryllidaceae* are separated from *Liliaceae* by their epigynous flowers. Hutchinson<sup>17</sup>, however, believes that the most essential character of true amaryllidaceous plants is their umbellate inflorescence subtended by involucral bracts. This results, *e.g.*, in transferring *Allium* and allied genera, which all possess hypogynous flowers from *Liliaceae* to *Amaryllidaceae*. The delimitation of the two families was rather profoundly altered by the new concept which has been accepted by several modern taxonomists and rejected by others. The question rises, which of the two concepts results in a more natural delimitation of the two families. In such instances chemical characters may aid taxonomists in finding the best answer. With regard to the example mentioned, present-day chemical evidence favours the traditional delimitation of *Liliaceae* and *Amaryllidaceae* with respect to *Allium* and related genera because the highly characteristic alkaloids of all true amaryllidaceous plants are seemingly lacking in the *Allium* alliance and because steroid sapogenins so wide-spread in *Liliaceae*, but apparently lacking in true *Amaryllidaceae*, do occur in *Allium* and allied genera. It is interesting to note that plant rusts seem to hold the same opinion; species attacking *Asparagus*, a liliaceous plant, attack also *Allium* but seem not to attack amaryllidaceous plants<sup>33</sup>. This, however, may not be an independent piece of evidence since host preference of parasites may largely be governed by the chemistry of the hosts' tissue.

### **Chemical characters as aids in unambiguous identifications of plants**

Plant species are composed of interbreeding populations of individuals. If a species has been highly successful and covers a large area at present, many of its populations become geographically and (or) ecologically separated. Gradually the gene pools of radiating populations may change and distinct topotypes or ecotypes may emerge. The latter may still be interfertile with all other populations of the species and clearly represent only variants of one wide-spread species. If, however, by polyploidy or some other mechanism barriers to gene exchange between the diverging entities have arisen or if clearcut morphological differences have evolved the matter of species delimitation becomes a delicate and difficult task. Many of the so called species aggregates have been taxonomically interpreted in different ways and nomenclature has often become complex and rather disappointing in such notoriously difficult groups. In this field of taxonomy cytological research has proved to be often successful. It may be an invaluable aid for an unambiguous identification of distinct entities and in many instances it has offered even a clue for a better understanding of the past history of such puzzlingly complex aggregates. Frequently past history gave rise to slightly differing metabolic patterns in members of a species aggregate. The study of their chemical constituents may therefore bring to light new

characteristics helpful in identification. It is evident that each botanical study concerning present-day distribution, ecological preferences and past history of members of an aggregate species or of several closely related species of a genus depends on the unambiguous identification of each available specimen. Unfortunately morphological characters are often rather vague and cytological work is restricted to living plants. Moreover clearcut distinctive morphological characters may be restricted to organs many often lacking in the available plant specimens. *Nasturtium officinale* R.Br. and *Nasturtium microphyllum* (Boenningh.) Rchb., for instance, can only be identified with certainty if mature fruits and seeds are present and the three sub-species of *Sparganium erectum* L. are identifiable by their fruits only. A thorough study of the chemistry of each member of such aggregates and the elaboration of analytical methods which may be performed even with herbarium specimens can be, in many instances, useful to plant taxonomy. I like to illustrate this aspect by an example, with which I have some personal experience. The species of ferns generally known as *Dryopteris filix-mas* (L.) Schott is a rather complex aggregate. The cytogenetical work of Manton<sup>30</sup> has shown it to comprise essentially three well defined entities in Europe, a fertile diploid called *Dryopteris abbreviata* Lamk. et DC., a fertile tetraploid called *Dryopteris filix-mas* (L.) Schott sensu stricto and an apogamous diploid or triploid called *Dryopteris borreri* Newm. Since several years we have investigated European species of *Dryopteris* for the phenolic compounds present in their rhizomes. Aided by Dr. J. Sundman and Miss A. Penttilä of Helsinki, who have studied intensively the chemistry of fern phloroglucides during recent years, we were able to show that the three members of the *filix-mas* aggregate differ distinctly in the composition of their phloroglucides. If adequately collected herbarium specimens are available these chemical characters can be very helpful in an unambiguous identification of dried plants, which have lost part of their morphologically most distinctive features.

There are, however, many more aspects, which make the study of chemical characters at infraspecific and specific levels a very fascinating one. Besides being helpful with the identification of plant specimens it informs us about patterns of chemical variation within genera and aggregate species and it may ultimately demonstrate how one pattern of plant constituents evolved from a preceding one. Moreover, joint botanical and phytochemical studies may provide us with a better understanding of the biological and ecological meaning of distinct spectra of primary and secondary plant metabolites. A thorough knowledge in these fields is essential for a judgement of the overall taxonomic implications of the overwhelming multitude of phytochemical patterns.

### SOME FACTORS WHICH LIMIT THE TAXONOMIC VALUE OF CHEMICAL CHARACTERS

To make the most appropriate use of chemical characters in plant taxonomy one has to realize clearly that several factors affect and restrict their taxonomic meaning. I would like to discuss especially parallelism and diversification and methods of documentation.

## CHEMICAL CHARACTERS IN PLANT TAXONOMY

### Parallelism and diversification

Every taxonomist is aware of the fact that morphological similarity of plants does not always indicate close relationship and that, on the other hand, striking dissimilarities can often be noted between taxa supposed to be closely related. These phenomena known as parallelism (convergence) and diversification (divergence) are very often responsible for difficulties and artificialities in classification. For taxonomists who endeavour to construct a natural system of plants it is most essential to analyze carefully each instance of suspected parallelism or diversification. As these phenomena often affect morphological characters it can be taken for granted that the same holds good for chemical characters. They too are in need of a careful analysis. First of all we should be able to discern true convergence from cases of pure analogy and true divergence from cases of clearcut homology. The already mentioned plant chromogenes arbutin and aucubin may serve to illustrate these points.

At present arbutin is known to occur in a number of plant families, some of which are distinctly not closely related to the other ones (*Table 3*).

*Table 3.* Some of the plant families in which arbutin has been found to occur

<i>Family</i>	<i>Remarks</i>
Ericaceae	Possibly remotely related to the rosaceous stock
Leguminosae	
Rosaceae	Families related by descent ("Rosaceous stock")
Saxifragaceae	
Rubiaceae	Possible remotely related to the rosaceous stock
Proteaceae	
Liliaceae	{ Distinctly non-related with the above-mentioned families

We need information concerning the biogenetical pathways giving rise to arbutin in *each* taxon known to accumulate this glucoside. If the pathway is the same in different taxa which are definitely non-related by a number of other criteria, then we have a clearcut example of convergence, i.e. convergent evolution. If, however, the pathway is a different one, arbutin accumulation in those taxa using a deviating pathway becomes a case of analogy; apparently the character is the same but it is acquired along different lines. It is obvious that analogous characters are of no value as guides for classification and that true convergence represents one of the many factors which make so difficult the design of a truly natural system of plants. At the same time we note that chemical parallelism may be achieved by two different processes with different taxonomic implications.

Aucubin and many closely related iridoid glycosides are especially widespread in *Saxifragaceae*, *Cornaceae*, *Garryaceae*, *Ericaceae*, *Oleales*, *Gentianales*, *Tubiflorae*, *Plantaginales* and *Dipsacales* (orders according to Englers' Syllabus). In some members of the taxa mentioned radically different

compounds replace aucubinlike glycosides. *Gentianaceae* produce gentiopicrin, swertiamarin and gentianin. The same is true for one tribe of *Loganiaceae*. *Oleaceae* accumulate oleuropein and many *Loganiaceae*, *Apocynaceae* and *Rubiaceae* synthesize the so called complex indolic alkaloids. Some members of *Apocynaceae*, *Bignoniaceae* and *Valerianaceae* are known to contain

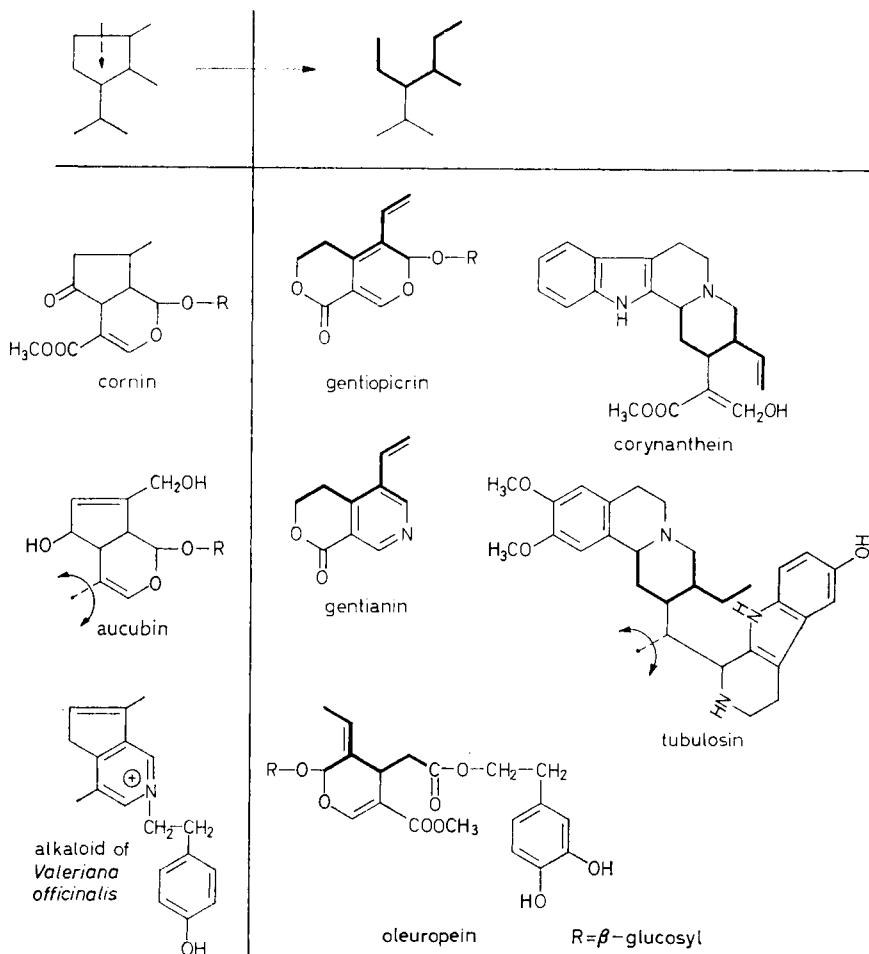


Figure 1. The homologous series of monoterpenoid iridoid, phytoconstituents

skyttanthin-like alkaloids. An overwhelming number of chemically unlike constituents (cf. Figure 1) has been detected in the taxa mentioned. All these compounds, however, according to a hypothesis of Thomas<sup>34</sup> could arise along biogenetically very similar lines. Their molecules would represent partly (indolic alkaloids) or wholly (skyttanthin-like alkaloids, gentianin, iridoid compounds) modified cyclopentanoid monoterpenes. Results of most recent investigations about the biosynthesis of different members of this assembly of phytoconstituents tend to confirm the hypothesis of Thomas.

## CHEMICAL CHARACTERS IN PLANT TAXONOMY

If this is the truth, the whole group of compounds represents a homologous series because the different members of the group are elaborated along essentially similar lines. Notwithstanding pronounced chemical dissimilarities of members of a homologous series of chemical compounds the latter may indicate true relationship of plants accumulating them. On the other hand if within a sharply defined genus such as *Pinus*, some members (e.g. *Pinus sabiniana* Dougl. and *Pinus jeffreyi* Balfour) produce predominantly *n*-hexane and pinidin as volatile constituents of their leaves instead of the usual monoterpenes, this represents a case of true divergence. The deviating compounds arise from another pathway. Like true convergence, true divergence may be a factor rendering very difficult the elaboration of a natural system of plants. Like parallelism, chemical diversification may originate in different manners and its bearing on taxonomic problems can only be evaluated after a far-reaching analysis of the underlying facts.

The following discussion will be devoted to parallelism only. Moreover, for convenience, three types of convergent evolution will be discerned.

### *Parallel overall evolutionary trends in phylogenetically remote taxa*

Biologically governed tendencies of flower and inflorescence evolution are rather well understood. Zygomorphic flowers or pseudanthia (a pseudanthium is a showy inflorescence imitating a single flower; all composites, e.g., bear pseudanthia) evolved independently in many insect-pollinated plant groups and inconspicuous, unisexual or protogynous flowers often aggregated in spiklet- or catkin-like inflorescences evolved in plants which reverted to wind pollination. Nobody classifies, e.g., all pseudanthia-bearing plants together because the individual flowers in pseudanthia usually preserve their characters and because some aspects of flower and inflorescence evolution resulting in many types of parallelism are rather well understood. For metabolic patterns and individual categories of constituents of angiosperms general tendencies of evolution are scarcely known at present. Alkaloids, for instance, have been detected in Fungi, Pteridophyta, Gymnospermae and Angiospermae but it is virtually impossible to indicate evolutionary trends concerning their structure and distribution. Within taxa of lower rank like species in a genus, genera in a family and even families in an order such tendencies may emerge in the near future but with regard to the whole plant kingdom such tendencies seem not to exist at all or are still far from being conceived clearly. Many instances of parallelism (e.g. distribution of senecionin-type pyrrolizidin alkaloids; distribution of aporphine-type alkaloids; distribution of tropan-type alkaloids) are known with regard to alkaloids but they are not yet understood in a satisfactory manner.

In other fields of plant chemistry the first indications for overall evolutionary trends begin to become apparent. Bate-Smith<sup>35</sup> and Lebreton<sup>36</sup> have put forward the hypothesis that in angiosperms, plants accumulating leucoanthocyanins and trihydroxylated phenols (leucodelphinidin, myricetin, gallic and ellagic acids) in leaves preceded plants not producing such compounds and that flavonol synthesis preceded the production of other types of anthoxanthins. If such general tendencies for the evolution of

phenolic patterns of leaves do indeed exist in angiosperms, many cases of chemical convergence (*e.g.* accumulation of apigenin and luteolin in non-related plant groups) become easily understandable.

Other examples of rapidly accumulating evidence for general evolutionary trends for plant metabolites may be found in the fields of the chemistry of lignin, hemicelluloses and cuticles and their waxes<sup>37</sup> and perhaps even in the field of triterpene chemistry<sup>38</sup>.

An insight in general evolutionary trends for categories of plant constituents (*i.e.* knowledge of their "Merkmalsphylogenie") is taxonomically important in other respects too. Taxa possessing many progressive characters can be derived from taxa with characters of a lower evolutionary level but the reverse, of course, is impossible.

#### *Parallelism arisen in connection with adaptation to environment*

Many morphological and anatomical characters of plants are intimately connected with adaptations to special exigencies of habitats. This immediately explains many cases of parallelism (and diversification) and prevents us from overrating the taxonomic implications of clearly adaptive characters.

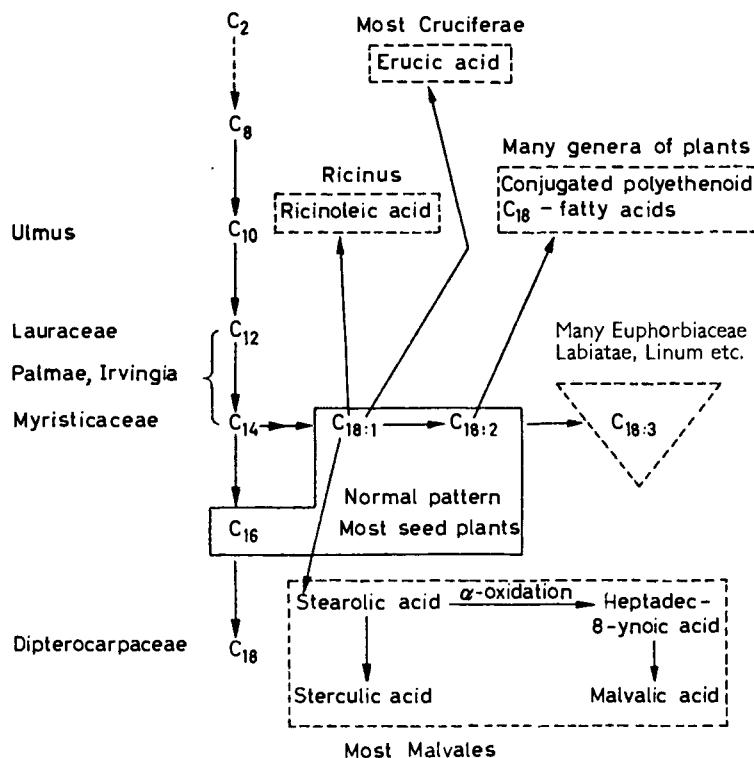
Most probably the accumulation of many of the highly curious secondary plant metabolites as well as distinct patterns of regular plant constituents are the result of selection by environment. It would not seem surprising that a metabolic variant perfectly adapted to fixed conditions of plant life originated more than once in phylogenetically non-related taxa. But to understand and interpret the facts correctly, it is essential to know something about the ecological and biological meaning of chemical patterns of plants. With regard to the majority of the so-called secondary plant constituents our knowledge in this field is extremely poor. Fraenkel<sup>39</sup> has gone so far as to declare that "these odd chemicals arose as a means of protecting plants from insects and now guide insects to food". This, of course, can be only part of the story. Climatic and edaphic factors of plant habitats are not less important as selecting agents than are insects. Their influence on phytochemical patterns, however, is scarcely known. Hillis<sup>40</sup> obtained indications that in eucalypts stilbene production is correlated with the aridity of habitats and it seems that in essential oil bearing plants many of the intraspecific chemotypes represent probably populations selected predominantly by microclimates. It has been suggested that the flavonoid persicarin occurs essentially in marsh plants<sup>41</sup> and if this proves to be true the compound may in somehow be involved in hygrophilic adaptation. Secondary plant constituents moreover may help some species in their competition with other plants for a given habitat. It must, however, be agreed that we are still unaware of the true contribution of secondary plant metabolites to the overall fitness for life of plants.

In the field of ordinary plant metabolites our present-day position is perhaps a little better.

Carbohydrate accumulation in storage organs of perennial plants is a common feature. In chlorophytes, bryophytes, pteridophytes, gymnosperms and angiosperms sucrose and starch generally fulfil a storage function. There exist, however, many groups of plants which have largely replaced sucrose and starch by other carbohydrates. It is highly probable that such replace-

## CHEMICAL CHARACTERS IN PLANT TAXONOMY

ments represent examples of progressive evolution correlated in some way with ecological features, e.g., with the colonization of less favourable habitats. Many rhizomatous species of *Iris* replaced starch partially or totally by irisin-type fructanes as they moved farther away from the Mediterranean centre of the genus. The Eurasian *Iris pseudacorus* L. and the Northern American *Iris virginica* L. and *Iris versicolor* L. store exclusively fructanes in their rhizomes. Similarly many of our perennial grasses are known to store fructanes in their rootstocks whereas tropical grasses seem to



*Figure 2.* Possible derivation of some characteristic seed oils of plants from the most common pattern<sup>44-48</sup>.

accumulate sucrose and starch preferentially. The storage sugars of our perennial *Labiatae* are stachyose and higher oligogalactosides of sucrose while some tropical members of the family have retained starch accumulation. In the subfamily *Silenoideae* of *Caryophyllaceae* starch is replaced by oligogalactosides formerly called lactosin and presently known as belonging to the lychnose and isolychnose group of oligosaccharides.

Fatty oil is the main storage product in the seeds of many flowering plants. Triglycerides containing palmitic, oleic and linoleic acids as main fatty acids are by far the most common ones. However, in the field of seed oils too many deviations from this common pattern are known (cf. Hilditch<sup>42</sup>; Shoreland<sup>43</sup>). Such aberrant seed oils may be looked upon as specializations,

i.e. as the result of a process of progressive evolution, which, in many instances, may have been governed by external factors. Recent observations and speculations suggest that most of the "unusual" fatty acids encountered in seed oils arise from oleic or linoleic acid, i.e. by the addition of new steps to the ordinary pathway of fatty acid synthesis in seeds. Other "unusual" seed oils may be derived from the "usual" ones by suppression of a few of the ordinary steps. Some of the suggested connections are illustrated in *Figure 2*, which is based on the scheme of James *et al.*<sup>44</sup>.

Several facts seem indeed to indicate that "unusual" seed oils originated in connection with ecological specialization. The predominantly tropical *Capparidaceae* produce seed oils of the normal type; the closely related extratropical *Cruciferae* have erucic acid as a main fatty acid in the seed oils of many of their members. Many *Labiatae* produce seed oils, rich in linolenic acid, while most of their tropical relatives seem to have oils of the ordinary type. In the genus *Cucurbita* trichosanic acid seems to be restricted to the specialized and genetically and geographically isolated xerophytic species *Cucurbita digitata* Gray, *C. palmata* Wats. and *C. foetidissima* H.B. et K. In this respect it is interesting to note that Rehm<sup>49</sup> reported that seedlings of the more advanced species of *Cucurbita* contain cucurbitacin E whereas seedlings of the more primitive species contain cucurbitacin B only. He investigated *Cucurbita palmata* and *C. foetidissima*, too; in seedlings of both species cucurbitacin E is present.

Distinct types of specialization are likely to have occurred in several non-related taxa. The more a special chemical character is connected with adaptation the more the incidence of parallelism is to be expected. If, for instance, we accept for the evolution of seed oils the progression oleic → linoleic → linolenic acid and if at the same time we are able to demonstrate clearly an advantage of the linolenic type in a cold climate then the fact that seed oils rich in linolenic acid are characteristic for several non-related taxa will no longer be an argument against the taxonomic potentialities of chemical characters.

#### *Accidental Parallelism*

Morphological parallelism seems to be purely accidental in many instances. It may solely be the result of an unlimited bias of nature for variation giving rise to a tremendous series of forms many of which may be neither profitable nor deleterious in the struggle for life and therefore will hardly be affected by selection through environment. One of the astonishing aspects of nature is its power to achieve a certain goal by a seemingly unlimited number of variants. Nature has been compared with a playing child<sup>50</sup> whose activities are not governed by economics and expediency but rather by imagination and by the pleasure in experimentation.

The number of non-related plants bearing similar leaves is very large. *Tropaeolum*, *Umbilicus* and *Hydrocotyle* or *Trifolium* and *Oxalis* may be cited as examples. Many types of chemical parallelism originate probably in a similar manner (*e.g.* isoflavones in *Podocarpus* and several families of angiosperms; bioflavonoids in *Coniferopsida*, *Casuarinaceae* and *Caprifoliaceae*).

### **Limitations caused by incorrect identification and by the omission of documentation**

Some of the preceding discussion should already have demonstrated that in many instances the correct identification of plant samples is a far from easy task. To sum up some factors causing difficulties and ambiguities the following ones very often may be involved.

1. Many species of plants are complex aggregates, their members being often characterized predominantly cytologically or ecologically. Their taxonomic treatment may change with time and may moreover be dependent on the systematists' personalities working at a given time with the aggregate. Taxonomy and nomenclature very often become highly troublesome and disappointing in such entities.
2. In many aggregate species and in many genera with taxonomically good but morphologically rather concealed species a correct identification implies a rather intimate acquaintance with the plants concerned.
3. Hybrid origin of plant samples may often cause difficulties of identification. Conditions for hybridization are especially favourable in Botanical Gardens where many species are grown close together.
4. Many floras of the world, especially those of tropical countries are still poorly known and only superficially studied from a taxonomic point of view. Every modern revision results in a large number of reductions of species and sometimes even genera. Recombinations and descriptions of new taxa are moreover considered necessary by every taxonomist revising a group of tropical plants or monographing a tropical genus or family.

The facts mentioned and many others imply that in many instances the result of plant identification depends on the paper or the flora used for this purpose. Even if a taxonomist is consulted for help with plant identification the name given to the plant material will be dependent on his acquaintance with and his personal interpretation of the respective group of plants.

There is only one means of escaping all ambiguities in the matter of plant identification. It consists in documentation. Each scientist working with plant material should understand and accept the obligation to document botanically his plant sources. This implies that perfect herbarium specimens are prepared and adequately (time and locality of collection) labelled. The specimens should be deposited in a herbarium accessible to other scientists. The specimen numbers and the institution where specimens were deposited should always be given in phytochemical publications. If herbarium specimens cannot be prepared because crude drugs (woods, seeds, commercial crude drugs) are investigated it should never be forgotten to preserve an adequately labelled representative sample of this material and to deposit it in a crude drug collection accessible to other scientists. To illustrate the importance of such a procedure I should like to give a recent example. Indian workers<sup>51</sup> isolated a series of coumarins from roots of *Nardostachys jatamansi* DC. (*Valerianaceae*) commercially available. These coumarins seemed out of place to me in *Valerianaceae*. Professor Bhattacharyya (of National Chemical Laboratory, Poona, India) was kind enough to send me a sample of the crude drug investigated. The anatomy of the roots indicated

R. HEGNAUER

clearly that the crude drug did not represent the rootstock of *Nardostachys jatamansi* but of an umbelliferous substitute.

An adequate documentation of the starting material of each phytochemical investigation is the only means of minimizing the consequences of the very frequent errors in plant identification, because documentation makes possible rechecking determinations at any given time. Phytochemical literature is full of errors of plant identification and plant naming. To prevent continuation of this undesirable situation every chemist and botanist working with plants should undertake the necessary steps, troublesome as they may be, to guarantee an adequate documentation of his starting materials. If one realizes that systematic botany is even more interested in the results of phytochemical research than chemistry, which can study its problems with pure synthetics as well, one will immediately perceive that the troubles involved in an adequate documentation will be recompensed by imparting a more general scientific value to the results of the investigations.

Greshoff<sup>52</sup> addressed the following words to an audience of scientists in 1890 "Wellicht gelukt het later deze hypothese (i.e. that chemical characters are valuable for taxonomic botany) meer zekerheid te geven, en komt de tijd dat de chemie op haar beurt aan de botanie een deel der goede diensten terugbetaalt, die deze wetenschap nu aan haar bewijst"?†. This period has now been reached without doubt. But phytochemistry must observe a meticulous documentation if she really intends to repay systematic botany part of the support which the latter science has always offered to her.

## References

- <sup>1</sup> A. P. DeCandolle. *Essai sur les propriétés médicales des plantes, comparées avec leurs formes extérieures et leur classification naturelle*, first edition, Paris (1804); second edition, Paris (1816).
- <sup>2</sup> M. Greshoff. *Ber. Deut. Pharm. Ges.* **3**, 191 (1893).
- <sup>3</sup> F. Rochleder. cf. e.g. *Ann. Chem. Pharm.* **83**, 64 (1852).
- <sup>4</sup> M. Greshoff. *Mededeelingen uit 's Lands Plantentuin No. VII* (Batavia 1890) and No. XXV (Batavia 1898).
- <sup>5</sup> L. Rosenthaler. *Beih. Botan. Zentralbl.*, 1. Abt. **21**, 304 (1907).
- <sup>6</sup> R. T. Baker, and H. G. Smith. *A Research on the Eucalypts and their essential oils*, second edition, Government Printer, Sydney (1920).
- <sup>7</sup> M. Wheldale. *Biochem. J.* **5**, 445 (1911).
- <sup>8</sup> S. Iwanow. *Beih. Botan. Zentralbl.*, 1. Abt. **32**, 66 (1915); *Ber. Deut. Botan. Ges.* **44**, 31 (1926).
- <sup>9</sup> H. Colin. *Rev. Gen. Sci. Pures et Appl.* **46**, 165 (1935).
- <sup>10</sup> H. Molisch. *Pflanzenchemie und Pflanzenverwandtschaften*, G. Fischer, Jena (1933).
- <sup>11</sup> J. B. McNair. *Studies in plant chemistry including chemical taxonomy, ontogeny, phylogeny, etc.* (A collection in book form of papers of the author between 1916 and 1945). Published by the author, Los Angeles, California (1965).
- <sup>12</sup> Th. Weevers. *Rec. Trav. Botan. Néerl.* **30**, 336 (1933); Koninkl. Akad. Wetensch. Amsterdam, Proceedings **39**, 757 (1936); *Blumea* **5**, 412 (1943).
- <sup>13</sup> M. Bridel, and A. Kramer. *Compt. Rend.* **193**, 748 (1931).
- <sup>14</sup> R. Wettstein. *Handbuch der systematischen Botanik*, 4. Aufl., F. Deuticke, Leipzig und Wien (1935).
- <sup>15</sup> A. A. Pulle. *Compendium van de terminologie, nomenclatuur en systematiek der zaadplanten*, third edition, N.V.A. Oosthoeks Uitg. Mij., Utrecht (1952).

† Perhaps it will be possible at a later date to judge better the taxonomic meaning of chemical characters and the time may arrive when chemistry will be able to render to botany part of the help the latter has always offered to phytochemistry.

## CHEMICAL CHARACTERS IN PLANT TAXONOMY

- <sup>16</sup> A. Cronquist. *Bull. Jardin Botan. Bruxelles* **27**, 13 (1955).
- <sup>17</sup> J. Hutchinson. *The Families of flowering plants*, second edition, Clarendon Press, Oxford (1959).
- <sup>18</sup> A. Takhtajan. *Die Evolution der Angiospermen*, G. Fischer, Jena (1959).
- <sup>19</sup> A. Engler. *Syllabus der Pflanzenfamilien*, 12. Aufl., Band 2, Gebr. Bornträger, Berlin (1964).
- <sup>20</sup> R. F. Thorne. *The American Naturalist* **97**, 298 (1963).
- <sup>21</sup> L. Benson. *Plant taxonomy. Methods and principles*. The Ronald Press Comp., New York (1962).
- <sup>22</sup> P. H. Davis, and V. H. Heywood. *Principles of Angiosperm taxonomy*, Oliver and Boyd, Edinburgh-London (1963).
- <sup>23</sup> H. Solereder. *Systematic anatomy of the Dicotyledons*, Clarendon Press, Oxford (1908).
- <sup>24</sup> C. R. Metcalfe, and L. Chalk. *Anatomy of the Dicotyledons*, Clarendon Press, Oxford (1950).
- <sup>25</sup> S. Carlquist. *Comparative plant anatomy*, Holt, Rinehart and Winston, New York (1962).
- <sup>26</sup> P. Maheshwari. *An introduction to the embryology of Angiosperms*, McGraw-Hill Book Co., New York-Toronto-London (1950).
- <sup>27</sup> R. P. Wodehouse. *Pollen Grains*, Hafner Publish. Co., New York (1959).
- <sup>28</sup> G. Erdtman. *Pollen morphology and plant taxonomy. Angiosperms*. The Chronica Botanica, Waltham, Mass. (1952).
- <sup>29</sup> C. D. Darlington. *Chromosome botany and the origins of cultivated plants*, second edition, George Allen and Unwin Ltd., London (1963).
- <sup>30</sup> I. Manton. *Problems of cytology and evolution in the Pteridophyta*, University Press, Cambridge (1950).
- <sup>31</sup> R. Hegnauer. Chemismus und systematische Stellung der Cornaceae, in *Festschrift Kurt Mothes zum 65. Geburtstag*, p. 235–246, G. Fischer, Jena (1965).
- <sup>32</sup> R. Hegnauer. *Chemotaxonomie der Pflanzen*, Vol. 3, p. 565–569 and fig. 29 on p. 544, Birkhäuser, Basel (1964).
- <sup>33</sup> D. B. O. Savile. *Science* **120**, 583 (1954).
- <sup>34</sup> R. Thomas. *Tetrahedron Letters* 544 (1961).
- <sup>35</sup> E. C. Bate-Smith. *J. Linn. Soc. London, Botany* **58**, 371 (1962).
- <sup>36</sup> Ph. Lebreton. *Bull. Soc. Botan. France* **111**, 80 (1964).
- <sup>37</sup> Cf. R. Hegnauer, *Chemotaxonomie der Pflanzen*, Band 1, p. 192–219 and 307–310, Birkhäuser, Basel, 1962; E. A. Baker and J. T. Martin, *Nature* **199**, 1268 (1963); A. M. Silva Fernandes, E. A. Baker, and J. T. Martin, *Ann. Appl. Biol.* **53**, 43 (1964); T. E. Timell. *Svensk Papperstidning* **65**, 266 (1962).
- <sup>38</sup> R. Hegnauer. *Lloydia* **28**, 267 (1965).
- <sup>39</sup> G. S. Fraenkel. *Science* **129**, 1466 (1959).
- <sup>40</sup> W. E. Hillis, and K. Isoi. *Phytochemistry* **4**, 541 (1965).
- <sup>41</sup> H. Tatsuta, and Y. Ochiai. *Sci. Reports Tohoku Univ.*, First series **39**, 236 (1956).
- <sup>42</sup> T. P. Hilditch, and P. N. Williams. *The chemical constitution of natural fats*, fourth edition, Chapman-Hall, London (1964).
- <sup>43</sup> F. B. Shoreland. The distribution of fatty acids in plant lipids, in T. Swain. *Chemical plant taxonomy*, p. 253–303, Academic Press, London-New York (1963).
- <sup>44</sup> A. T. James, R. V. Harris, C. Hitchcock, B. J. B. Wood, and B. W. Nichols, *Fette, Seifen, Anstrichmittel* **67**, 393 (1965).
- <sup>45</sup> G. N. Smith, and J. D. Bu'Lock. *Chem. and Ind.* 1840 (1965).
- <sup>46</sup> F. D. Gustone. *Chem. and Ind.* 1033 (1965).
- <sup>47</sup> D. T. Canvin. *Canad. J. Botany* **43**, 49, 63, 71 (1965).
- <sup>48</sup> Z. B. Kondra, and B. R. Stefansson. *Canad. J. Genetics Cytol.* **7**, 500 (1965).
- <sup>49</sup> S. Rehm. *Ergebnisse der Biologie* **22**, 108 (1960).
- <sup>50</sup> K. Mothes. *Naturwissenschaften* **52**, 571 (1965).
- <sup>51</sup> S. N. Shanbhag, C. K. Mesta, M. L. Maheshwari, S. K. Paknikar, and S. C. Bhattacharyya. *Tetrahedron* **20**, 2605 (1964); **21**, 3591 (1965).
- <sup>52</sup> M. Greshoff. Planten en plantenstoffen; lecture held on 11 December 1890 in Batavia; published by Koninklijke Natuurkundige Vereniging in Nederlands-Indië, (voordrachten No. 7), Ernst and Co., Batavia and Noordwijk (1891).

# PLANT CHEMOSYSTEMATICS AND PHYLOGENY

B. L. TURNER

*The University of Texas, Austin, Texas, U.S.A.*

## INTRODUCTION

There has been a fervour of intellectual activity in botanical systematics during the past 10 years, most of this associated with new approaches to old problems. Constance<sup>1</sup> has presented an excellent summary of this activity in his account of "Systematic Botany—An Unending Synthesis", and it need only be noted here that he recognizes three major new approaches to present-day systematics, these being (1) Chemical, (2) Numerical, and (3) Ultrastructural ("Fine Structure"). (I prefer to include ultrastructure as a subdivision of the chemical<sup>2</sup>, but at the present state of its development, and on pragmatic grounds, it would seem better to treat this approach as a subdiscipline in its own right.)

In view of the current impact of the chemosystematic approach it is perhaps appropriate to evaluate the place of this discipline amongst those of anatomy, cytogenetics, etc. Alston and Turner<sup>3</sup> have outlined the history of systematics by recognizing five major periods, the inception of each being characterized by the development of new concepts or approaches which have permitted the accumulation of new data and/or conceptual insights bearing on systematics generally. Briefly these are: (1) the Megamorphic (400 B.C.—1700 A.D.), (2) the Micromorphic (1700–1860), (3) the Evolutionary (1860–1900), (4) the Cytogenetical (1900–1960), and (5) the Biochemical (1960–?)†. It should be obvious that each of these periods has made, and continues to make, important contributions to systematics; in fact it might be said that each has contributed in proportion to its time-span since ushered in (*Figure 1*). Certainly the development of succeeding periods has depended upon those preceding, and it is highly likely that any successful evaluation of the biochemical must depend upon the mega- and micro-morphical data at all taxonomic levels.

That the early 1960s has truly ushered in the Biochemical Era can be ascertained from the fact that four major chemosystematic texts appeared during this period<sup>3–6</sup> as well as several major review articles<sup>7–14</sup>. An additional catalytic factor in plant systematics has been the establishment of the periodical *Phytochemistry* (commenced, 1961). Most of the articles published in this journal have considerable bearing on the field of plant chemosystematics. Finally, it should be noted that an International Committee on Chemotaxonomy was formed during this period—a joint association of the formal Committees of Chemotaxonomy of the Organic Chemistry Division of

† Numerical taxonomy<sup>15</sup> does not rate a position in the hierarchy since, as indicated by Constance<sup>1</sup>, it furnishes no new data, and in concept it is pre-Darwinian.

B. L. TURNER

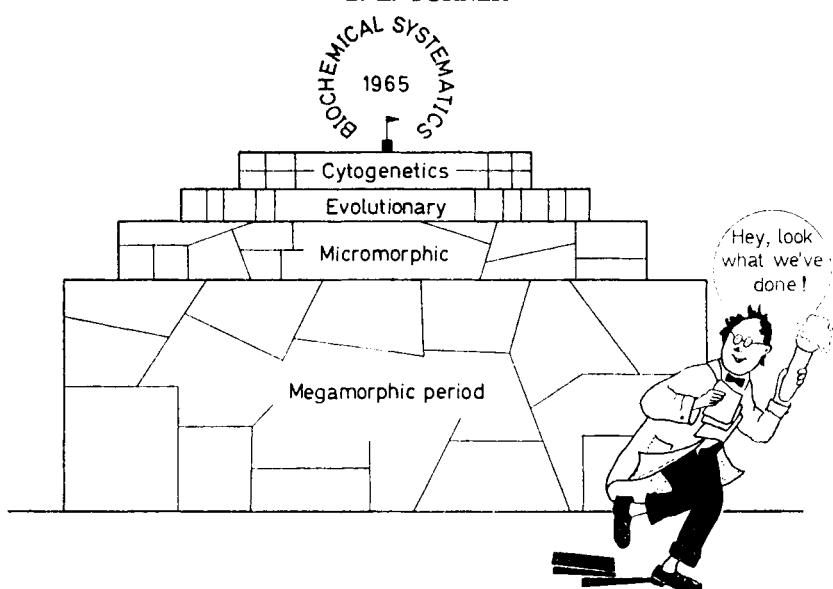


Figure 1. A chemosystematic perspective—the actual contribution of chemical data to the “taxonomic pyramid” in the year A.D. 1965. Compared to the megamorphic base, the chemical block at the apex has been exaggerated; additional explanation in text.

the International Union of Pure and Applied Chemistry and that of the International Association of Plant Taxonomists<sup>16</sup>.

Botanical chemosystematics did not develop full blown, for many early workers attempted to use the meagre data then available for taxonomic purposes<sup>17</sup>. Rather its emergence has coincided with the development of both rapid and relatively simple techniques for chemical isolations such as paper and gas-liquid chromatography. The chemosystematist for the first time could afford to undertake the broad samplings that are so necessary before meaningful generalizations can be made. As indicated by Erdtman<sup>18</sup>, most organic chemists are interested in nut cracking, not nut gathering. The chemosystematist surely must be interested in both aspects, hence the need for close collaboration between workers in these fields.

Such developments are not solely responsible for the surge of recent activity in chemosystematics, rather it is likely that enthusiastic acceptance, at least by taxonomists, has depended as much on its *demonstrated* contribution to the solution of specific systematic problems. Unfortunately, publications in which chemosystematics have helped solve taxonomic problems are relatively few, but their numbers are increasing rapidly and it seems likely that, as familiarity with the relatively simple chromatographic techniques increases, there will be a deluge of chemosystematically oriented articles, comparable to those ushered in by the Cytogenetical Period.

This accumulation of chemical data cannot be ignored by the plant taxonomist. The latter must become familiar with not only the molecular structure of compounds, but he should also possess some familiarity with biosynthetic pathways leading to their origin and, in general, appreciate the

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

kind of conceptual thinking that leads from triplicate sequences in the DNA molecules to amino acid sequences in proteins to the catalytic activity of the enzymes themselves. Yet it is true that the present day taxonomists are largely ignorant of these kinds of data. Their training has not emphasized a molecular approach to systematics and for the most part they have chosen to ignore the data already assembled, much less embark on selected studies of their own. Likewise, organic chemists who attempt to apply their data to systematic problems can gain better perspective if they become familiar with the conceptual basis of modern biological and systematic thinking, especially at the populational or evolutionary level. Other areas in which the chemist needs enlightenment have been covered by Professor Hegnauer in the present symposium. In fact, the negative attitude towards chemical data expressed by some plant taxonomists can be traced to the naïve taxonomic conclusions of chemists of an earlier period and persisting even today<sup>19</sup>, in spite of adequate admonitions<sup>8</sup> from chemists themselves. Hopefully this situation will change with the next generation thus permitting maximum interplay between systematics and organic chemistry. Indeed, it is already clear that natural plant product chemistry is being welcomed into biology departments, to the profit of both disciplines.

If chemistry assumes an increasing role in taxonomic thought and practice, one might well ask how molecular data can contribute most to systematic problems. Should these be applied towards the elucidation of evolutionary or phyletic problems or treated as unit data in a purely phenetic or numerical approach to classification? Chemical data possess a statistical objectivity<sup>20</sup> applicable to phyletic problems unobtainable from purely morphological studies (*e.g.*, the amino acid sequence studies of cytochrome *c*, haemoglobin, myoglobin, *etc.*<sup>18</sup>). In fact, the *unique* contribution that chemistry can make to systematics is that of providing a less than intuitive approach to the phyletic relationships of the higher categories generally. To reduce chemical data to mere listings in a purely numerical system would be getting the least out of the most. In order to appreciate this statement I would like to discuss, briefly, the two schools of thought, numerical taxonomy versus phylogeny, that currently hold forth among taxonomic circles.

## NUMERICAL vs. PHYLETIC CLASSIFICATION

It is a curious fact that, in spite of the almost universal acceptance of the evolutionary theory by biologists, there exists today a number of taxonomists who are opposed to<sup>15</sup> (or at least despair at the prospects of attempting<sup>21</sup>) classifying systems based on phyletic considerations. This school of workers would prefer to erect a so-called "natural" system in which essentially morphological data are quantified and computerized such that supposedly highly objective groupings might be made. The results may or may not reflect phyletic groupings and as a consequence this general approach to classification has been referred to as numerical taxonomy<sup>15</sup>. The pros and cons of the purely numerical versus phyletic approach to systematics has been amply expounded by experts in the area<sup>22-24</sup> and it need only be noted here that the phyleticists seem to have carried the day. At least Sokal, the principal spokesman for the phenetic (or numerical) approach, has recently

B. L. TURNER

turned his attention to numerical methods for the detection of "branching sequences in phylogeny" and only admits tacitly that a "phenetic basis is preferable for classification *in this narrow sense* [an arrangement of organized nature into categories for the convenience of biologists] until an *operational system*, combining cladistics [*i.e.*, phylogeny as used here] and phenetics can be established (italics mine)<sup>25</sup>. A view that most plant taxonomists would ascribe to, since this is exactly what they have been about for years; phenetic systems must necessarily serve as the initial model; the latter is translated into a hypothetical phyletic Model as new insights are gained, especially by the kind of *a posteriori* judgements referred to by Mayr<sup>26</sup> and through data which are difficult if not impossible to quantify such as that obtained from phytogeographical and cytogenetical considerations. Nevertheless, for the taxonomy of the plant kingdom, and in particular for that of the flowering plants, the numerical approach has held a certain attraction, mainly because fossil sequences are difficult to come by (the ultimate proof of phylogeny according to many workers) and the arrangement of the higher categories (*i.e.*, families and orders) *has* received a variety of differing treatments, in spite of the fact that each is purportedly phyletic in nature. Yet it is not the phyletic evidence that is the cause of the problem, rather it is the morphological complexity itself that has lead to such differing interpretations. It is unlikely then that purely numerical treatments, using morphological data, will resolve to any great extent the more vexing positional problems.

But, after all, is the chemist really interested in evaluating his data against hypothetical phyletic models? Perhaps he would be just as content to contribute these as unit data in a computerized numerical model<sup>26</sup>. To me the answer is clear enough, for it seems that most chemists are interested not only in the distribution of compounds, but also in their biosynthetic pathways, whether these pathways are advanced or primitive and where, in fact, existing pathways might have branched out of those pre-existing, etc. Indeed, on a character basis they are as phyletically minded as most evolutionary taxonomists<sup>27</sup>. Besides, these are intellectual matters. If evolution has been a fact and if the positional relationships of present day taxa reflect that process, then as scientists it is our responsibility to at least attempt to detect those arrangements.

In the discussion that follows then, the contributions of chemosystematics to plant taxonomy, both present and future, are taken to be primarily phyletic. These will be discussed under two general headings, Micromolecular and Macromolecular, depending on the relative molecular weight of the compounds under consideration. This seems a convenient way to organize current data since these two groups of compounds demand quite different techniques and often call for differing interpretations of otherwise comparable information†.

† Lanni<sup>11</sup> distinguishes only two kinds of taxonomy, *molecular* and *classical*: the former uses macromolecular data, either directly or by implication, to determine the base sequences of genetic material for comparative purposes; classical taxonomy using the remaining properties. Of course, such a rigid distinction would exclude micromolecular compounds. But molecules are molecules and perhaps Lanni would admit to the breakdown of molecular taxonomy into the two headings suggested here. Even so, there would be all shades of macromolecularism, grading from the smallest peptide through the aggregational macromolecules which make up the cellular organelles themselves.

## MICROMOLECULAR SYSTEMATICS

Most of the early chemosystematic literature, and much of the present, has concerned itself with the mere notation of the distribution of relatively simple compounds from only one or, less often, several collections of a given taxon. Among the more popular small molecular weight compounds investigated in this fashion have been those of amino acids, alkaloids, terpenoids, and flavonoids. These four chemical classes are widespread among plant groups but in each may be found certain subclasses which are restricted to closely related taxa. On the other hand, certain compounds in these groups occur in taxa which are clearly unrelated and it can only be assumed that their appearance in the organisms concerned reflect more the vagaries of the metabolic systems that produce them rather than they do phylogeny or gene homology. Some of these compounds, such as the 20 protein amino acids, are ubiquitous and hence of little or no value for taxonomic purposes. At the other extreme, one may point to the free amino acid lathyrine, which has been found in only 11 species of the genus *Lathyrus*, in spite of extensive surveys for this compound among plant groups generally<sup>28</sup>.

Like morphological characters, chemical characters are often variable, hence one of the first objectives in a chemosystematic survey is to establish the reliability of the data, both within and between populations. Too few of the early workers took such pains, and as a consequence, many premature systematic conclusions were drawn from incomplete data. As much care should be taken in acquiring chemical data as is taken for morphological measurements. Thus, for chromatographic surveys extracts for analyses should be made from comparable organs in comparable stages of development, preferably under comparable environmental conditions. This has been amply emphasized by Erdtman<sup>8</sup> and others.

Nevertheless the question which has been most often raised by workers in chemosystematics has been that of the effect of environmental variables on chromatographic patterns<sup>29</sup>. This question has been effectively answered for flavonoids by growing aseptically cultured, uniclonal-derived populations of *Spirodela oligorrhiza* and other aquatic species under a wide range of controlled environmental conditions<sup>30</sup>. In nearly all instances the flavonoid patterns of the taxa were unaffected by these variables; indeed the frond morphology was more often modified than were the chromatographic patterns, and identification in such instances could be more surely made by chromatography than by visual inspection of the fronds themselves. This is not to say that these compounds are never affected by such variables; as in morphology, the phenotypic expression of a given character should be reassessed as one undertakes the study of yet other taxa.

### Hybridization

Most chemists, in that they are not generally involved with purely systematic problems, are probably unaware of the widespread occurrence of natural hybridization among plant species generally, consequently they have perhaps not appreciated the significance of chemical data for resolving problems of natural hybridization. Morphological and cytogenetical

B. L. TURNER

techniques can be used to detect hybridization in most instances<sup>31</sup> but there are situations where chemical data can be decisive in dealing with such problems, particularly where complex hybridization is encountered. Thus, Alston and Turner<sup>32</sup>, when confronted with a natural hybrid swarm involving three species of the genus *Baptisia*, were able to show by two-dimensional paper chromatography that what appeared to be on morphological grounds, complex three-way hybridization and backcrossing was in

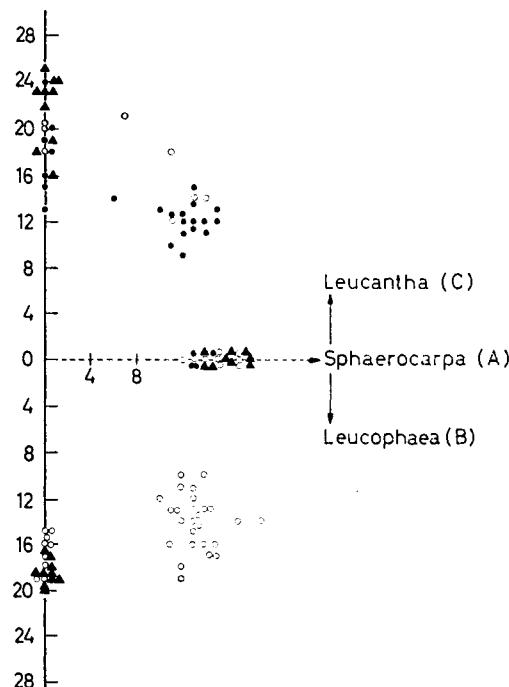


Figure 2. A three-way plot of individual hybrid types and pure species of *Baptisia*. Open circles indicate plants from a tri-hybrid population, closed triangles indicate miscellaneous supplementary plants from pure populations and closed circles indicate the additional (supplementary) *B. leucantha* × *B. sphaerocarpa* hybrids from other populations. Points along the horizontal (X) axis represent the number of compounds recognized of *B. sphaerocarpa*; points along the vertical (Y) axis represent (above) the number of compounds recognized of *B. leucantha* and (below) the number of compounds recognized of *B. leucophaea*. Hybrids fall at some angle between the X and Y axes (from Alston and Turner<sup>32</sup>)

reality a less complex situation in which crossing was mostly two-way: for example, the sympatric species A, B, and C were found in the following hybrid combinations, A × B (and backcrosses), A × C (and backcrosses) but the combination B × C (or backcrosses of the B genome into that of the C or vice versa), which was predicted on morphological grounds, could not be detected (Figure 2). An even more complex natural situation was reported<sup>33</sup> in which four species of *Baptisia* occurred together and, within a several acre field, all six of the possible two-way F<sub>1</sub> combinations were detected, along with several backcrosses. It is doubtful that these hybrid

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

populations could have been so successfully analysed without the aid of paper chromatography. One reason that the chemosystematic approach, using flavonoids in this instance, has been so successful in *Baptisia* is that, unlike morphological variables which are usually expressed through multiple gene systems<sup>34</sup>, the presence or absence of flavonoid components are usually governed by genes expressing dominance or recessiveness, the appearance of a compound generally reflecting dominance. Thus, the chromatographic pattern of an F<sub>1</sub> between two species of *Baptisia* with differing flavonoid patterns is expressed additively (*Figure 3*), i.e., the flavonoids found in the hybrid are the same as those found in *both* the parental types (analogous to chromosomes in amphiploids). Smith and Abashian<sup>35</sup> have reported the same phenomenon for alkaloids in *Nicotiana* hybrids, although several workers<sup>36,37</sup> have reported occasional synthetic hybrids where complementations of alkaloids were not found. However, when a large number of secondary substances are involved, it is unlikely that occasional variation in the anticipated patterns will prove grossly misleading. Thus, Parks<sup>38</sup> attempting to find whether striking differences in visual petal pigmentation in *Gossypium* might be due to relatively simple mutations which affect the entire spectrum of flavonoids, concluded "that flower-colour mutans which grossly affect visual colour of the flower do not obscure the basic array of flavonoids of the species studied, and that the species may be positively identified by the residual array of pigments not affected by particular alleles." In fact, McHale and Alston<sup>39</sup> working with selected species of *Baptisia* have suggested that for the detection of F<sub>1</sub>s, the chemical data may be more instructive than morphological data, for what they took to be unequivocal F<sub>1</sub>s on chromatographic grounds, appeared to be collections which morphologically strongly approached one or the other parental types.

One of the more spectacular studies of hybrids using chromatographic techniques has been that of Smith and Levin<sup>40</sup> for the genus *Asplenium*. These authors showed that the allopolyploid hybrid, *A. × kentuckiense*, which is believed to combine the genomes of three diploid species, has a chromatographic profile which is the additive combination of the distinctive profiles of each of the *Asplenium* species which contributed to the three-way hybrid. Their results agreed in all respects with the concept of reticulate evolution as proposed for this taxon by other workers, who used primarily morphology and cytology in their studies.

Hybrid studies of a similar nature, but with less striking documentation, have been made for species of *Viola*<sup>41</sup>; *Coprosma*<sup>42</sup>; *Zinnia*<sup>43</sup>; *Lotus*<sup>44</sup>; *Tragopogon*<sup>45</sup>; *Phlox*<sup>46</sup>; and *Vernonia*<sup>47</sup>. But at least a few workers have not found flavonoids particularly useful in the documentation of hybrids, thus Stromnaes and Garber<sup>48</sup>, working with *Collinsia*, could not detect F<sub>1</sub>s by this method even when the parental types possessed different chromatographic profiles.

Quantitative variation in chemical components, as opposed to qualitative, may also be used for the detection of hybridity as indicated by the excellent populational studies of Mirov<sup>49,50</sup>, Bannister *et al.*<sup>51</sup>, and Forde<sup>52</sup>. These workers have shown that F<sub>1</sub> hybrids between different species of *Pinus* possessed quantities of terpenes more or less intermediate to those possessed

B. L. TURNER

by the parental types, but they cautioned against the utilization of such data for the detection of  $F_2$  hybrids and backcross types since the quantitative variations, at least in part, seemed to be under complex genetic control. A similar study of a more restricted nature is that of Stone *et al.*<sup>53</sup> who used the quantitative variations of five fatty acids to help establish the hybrid nature of *Carya × lecontei*.

Aside from the documentation of hybrids *per se*, chromatographic data have provided a convenient means for the detection of *introgression*. As indicated by Anderson<sup>31</sup> the detection, by morphological criteria, of the influx of genes from one taxon into another by hybridization and back-crossing is most difficult, mainly because the inheritance of such characters is multifactorial and one or a few such genes, when incorporated into the variable gene-plasm of an adjacent organism, are difficult to pick up, for the environmental variables and the phenotypic plasticity of most plant populations are such as to preclude their discovery even by rather sophisticated statistical means.

When the species investigated possess quite different flavonoids, the relatively simple genetic control of such compounds<sup>54</sup> ought to make possible the detection of both allopatric and sympatric introgression. Horne<sup>55</sup> has suggested that certain regional variations in the flavonoids of *Baptisia nuttalliana* might be due to introgression; and Baetcke<sup>56</sup>, through his chromatography study of over 1400 plants of a hybrid swarm involving *Baptisia leucophaea* and *B. sphaerocarpa*, in my opinion, has demonstrated sympatric introgression between these species.

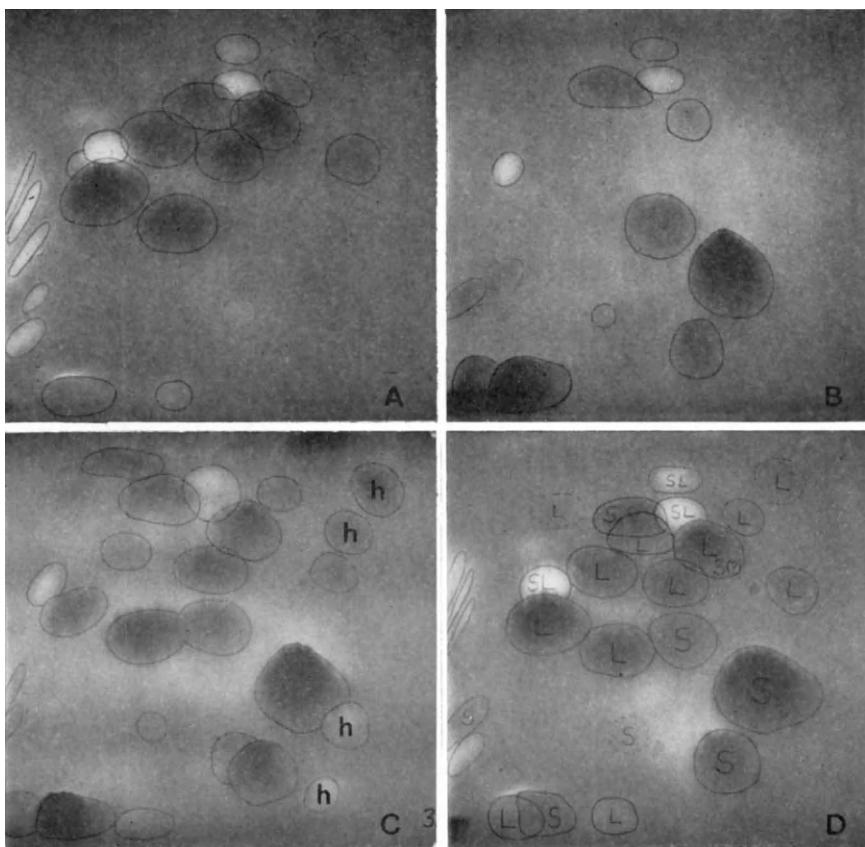
Gas chromatographic studies promise to be an important approach for the analysis of allopatric introgression (where two essentially geographically contiguous species exchange genes regionally as a result of usually long-time hybridization and backcrosses). The latter is difficult to detect by purely morphological studies, but the sensitivity of the gas chromatographic apparatus should make populational samples easy to quantify, especially where terpenoid compounds have been selected for examination. Von Rudloff<sup>57</sup> has been able to get reliable terpenoid data from a single conifer leaf and it should prove feasible to re-examine with chemical data<sup>58</sup> the now classical example of allopatric introgression in *Juniperus*.

The chromatographic method, at least for the detection of flavonoids, alkaloids, and most terpenoids, need not be limited to fresh material. Dried specimens as obtained by the usual herbarium procedures are often sufficient; in fact, using flavonoids, the hybrid nature of herbarium specimens over 85 years old or more may be established<sup>59</sup>. Finally, it should be noted that macromolecular data have also been used to document hybridization and several cases are noted in the section on Macromolecular Systematics (below).

### Phyletic groupings

In the discussion above, emphasis has been placed on the utility of chromatographic data for the detection of hybridization. This is one area of systematics where the chromatography of secondary compounds has made significant contributions to the phyletic process. Such an approach does not

PLANT CHEMOSYSTEMATICS AND PHYLOGENY

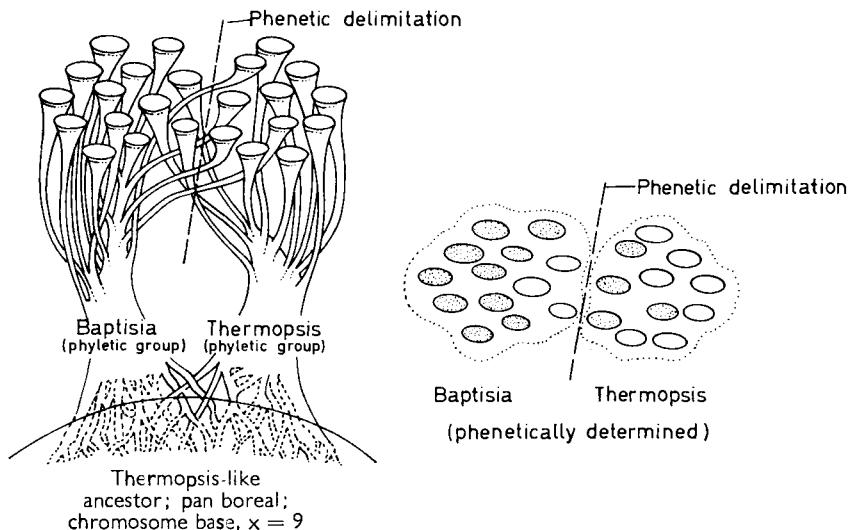


*Figure 3.* Chromatographic patterns of leaf extracts from *Baptisia leucantha* and *B. sphaerocarpa* and their  $F_1$  hybrid. A, *leucantha*; B, *B. sphaerocarpa*; C, natural hybrid; D, *in vitro* "hybrid" (obtained by mixing leaf extracts of the two species prior to chromatography). The letters L and S refer to the origin of the spot from either *B. leucantha* (L) or *B. sphaerocarpa* (S); the letter h refers to "hybrid spots" which appear in the natural hybrid (for biological reasons<sup>60</sup>) but do not occur in the *in vitro* "hybrid". Adopted from Alston and Hempel, *J. Hered.* **55**, 267 (1964)

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

require that the compounds be identified chemically, although such information might be desirable in attempting to evaluate the significance of any new compounds which might be detected in hybrids<sup>60</sup> (Figure 3).

While the relative differences of chromatographic profiles (e.g., general comparative studies of restricted groups of compounds such as flavonoids which are widely distributed among plants) might be of some utility in distinguishing between taxa<sup>61-63</sup>, the spot data, taken alone and without chemical determinations, are unlikely to contribute any real phyletic insights.



*Figure 4.* Hypothetical relationship of present day *Baptisia* and *Thermopsis* species. An evolutionary treatment would attempt to group the species into two monophyletic taxa (indicated by stippling and non-stippling). A purely numerical or phenetic treatment (*i.e.*, without phyletic considerations) might group the species into two polyphyletic genera

Thus one must distinguish between recognition, circumscription and arrangement. The former are only rarely affected by phyletic thinking; the latter often so. Phylogeny then, to most plant systematists, is the business of arranging taxa or groups of taxa such that their taxonomic position one to the other reflects best the genetic ancestry of those taxa. As indicated by Henning<sup>24</sup>, fossil records are not prerequisite for evolutionary groupings of extant taxa. To make this point, I would like to discuss briefly two possible approaches, phenetic vs. phyletic, to the arrangement of the species which comprise the genus *Baptisia*.

*Baptisia*, as currently treated, is composed of 16 easily recognized species, *the members of which are more closely related one to the other than any is to yet another taxon*. The taxonomic cliché indicated by italics serves as the conceptual "philosophy" of both phenetic and phyletic workers, with one important difference: the phyleticists would urge that the arrangement of relationships arrived at be made on evolutionary grounds; the pheneticists presumably would not bother to ask the question. *Figure 4* illustrates, albeit crudely, the problem for *Baptisia* and the closely related genus *Thermopsis*. On purely phenetic grounds, say by numerical means where all characters are given

B. L. TURNER

equal weight without *priori* considerations of any sort, the species which comprise these genera might well be grouped differently than they are at present. Indeed, when I first came across a rare species of *Thermopsis* in the south-eastern United States I thought I had discovered an undescribed species of *Baptisia*, so alike are these genera in vegetative and floral characters. I drew up a description of the supposed new species, but since the material lacked pods I returned to the population the following season and discovered that its fruit character was diagnostic of *Thermopsis*. This occasioned a re-examination of all my evidence, including the characters of *Thermopsis*, and I concluded that the population concerned was not a species of *Baptisia* with flat pods (a key character used to distinguish the genera; *Baptisia* possesses inflated, globose to cylindrical pods); rather it was a bona fide *Thermopsis*. Indeed, the species had already been described as belonging to that genus by an earlier worker. (I cannot help but add that on phenetic grounds I might have at least got a *name change* for my efforts; fortunately for nomenclature stability, my phyletic views prevailed.)

Quite apart from the morphological characters, there are other compelling reasons for treating *Thermopsis* and *Baptisia* as distinct phyletic groupings. For example, *Thermopsis* is a widespread boreal group occurring at high elevation (mostly in tiaga or timberline plant associations) from India across China into Russia, Alaska, western North America with an isolated series of species in the Appalachian Mountains of the eastern United States. Many of its species are isolated clonal relics, possessing little variability and only rarely showing evidence of evolutionary vigour. *Baptisia*, on the other hand, is restricted to the eastern United States and adjacent Canada, occurring at low elevations mostly in relatively warm moist regions, occupying floristic provinces which are believed to be younger than those occupied by *Thermopsis*. Its species are highly variable, exchange genes readily with other populational aggregates, reproduces readily from seed, increases rapidly in disturbed areas and, in general, acts as an evolutionarily aggressive taxon.

Phyletically speaking one might go further and suggest that a *Thermopsis*-like ancestor gave rise to *Baptisia*, perhaps from species of *Thermopsis* not too unlike those which persist today as relics in the Appalachian mountains. To obtain *Baptisia* one need only imagine an ancestral gene pool which, through successive changes, gave rise to populations with inflated pods, physiological adaptions to warm, moist habitats and whatever else it took to become successful in the new habitats which appeared following the post early Tertiary appearance of the coastal land masses of the southeastern United States. The evolution of *Baptisia*, as it is known today, was seemingly accompanied by an accelerated change in its vegetative features, the species possessing trifoliolate (presumably primitive) bifoliolate, unifoliolate and even perfoliate leaves. In addition, its pods have progressed from an erect, inflated, linear pod (presumably primitive) to a pendulous, globose, papery to woody pod. None the less, this remarkable proliferation in leaf and fruit types has not diminished the ability of these species to exchange genes through hybridization, for whenever two or more of the species grow together or near each other, hybrids can usually be found. (At those few sites in the Appalachian region where *Thermopsis* and *Baptisia* grow together, hybrids have not been detected.) In short, one simply cannot equate vegetative

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

characters in the two genera, for their evolutionary patterns, as described above, seem to have been different. Taking all the evidence, it is apparent that there are two distinct *phytic* series, *Baptisia* and *Thermopsis*; some species of *Baptisia* may look like *Thermopsis* and *vice versa*, but in my opinion, such resemblances are due either to parallelism from ancestral prototypes or to convergence.

It should prove instructive to evaluate the chemical data against the phyletic views expressed above. Unfortunately, *Thermopsis* has not been sufficiently investigated to permit meaningful evaluations. *Thermopsis* does possess a rich pool of flavonoids<sup>64</sup>, however, and a comparison of specific molecular types with those of *Baptisia* may permit some insight into the origin of the latter genus (*i.e.*, whether *Baptisia* is a monophyletic group arising out of a *Thermopsis*-like ancestral stock, relics of which still persist in the Appalachian Mountains; or whether *Baptisia* is polyphyletic, the extant species having arisen out of a number of *Thermopsis*-like ancestral lines, relics of which persist in the timberlines of Nepal, tiaga of Siberia and the mixed evergreen deciduous forest of the eastern United States).

The infrageneric distribution of flavonoids in *Baptisia* (Table 1) does suggest, however, that certain specific groupings are more likely than others. Interestingly enough, the chemical data support an alignment based on characters of the pod—an alignment which would have been difficult to detect using vegetative features as important criteria in the morphological evaluations. A possible exception is that of *B. megacarpa*. This species is a rarely encountered species of the Appalachicola River area of western Florida, a region renowned for its relic floristic elements. *B. megacarpa* is included with the white flowered species (*B. alba* and *B. leucantha*) on both its fruit and vegetative features (vegetatively like both *B. leucantha* and *B. alba*; fruit similar to *B. leucantha*) but floral characters suggest a relationship to the *B. leucophaea* series (Table 1), as do the flavonoids. In fact, it is likely that in placing the species with *B. alba* and *B. leucantha* we have been unduly impressed with the vegetative characters and perhaps superficial appearance of the fruit characters. In hindsight, we seem to have ignored the floral characters; but the recently assembled flavonoid data have forced us to take a second look and it might be that the *phytic* position of *B. megacarpa* is, after all, with the *B. leucophaea* alliance. At least the chemical data do suggest that it is an exceptional species, perhaps a relic from some ancestral complex which gave rise to both the *B. leucantha* and *B. leucophaea* groups; certainly it is not a recent populational derivative of the former species.

In addition to the flavonoids, at least 15 lupine alkaloids have been reported for *Baptisia* and *Thermopsis*<sup>65,66</sup>. Both genera possess the tricyclic compounds cytisine, methylcytisine and anagyrine as their major alkaloids and consequently these data have proven to have little systematic value at the generic level or lower. The alkaloids do however indicate that the two genera (and their northern hemisphere relatives *Piptanthus* and *Anagyris*) are perhaps rather distantly related to the tribe Podalyrieae (where they are usually placed), standing instead somewhere between the southern hemisphere elements of that tribe and the tribe Genisteae (Figure 5). In this connection it should be noted that *Thermopsis* possesses a larger pool of alkaloids than does *Baptisia*, many of these of the tetracyclic type

## B. L. TURNER

Table I. Flavonoids of *Baptisia*<sup>a</sup> (Leaf Extracts)

⊕ Compound present     ● Detected in only trace amounts

Species	Flavonols <sup>b</sup>												Isoflavones <sup>c</sup>																	
	Flavanones <sup>d</sup>						Dopag <sup>e</sup>						Flavanonols <sup>f</sup>						Flavanones <sup>g</sup>											
	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	10	11	1	2	3	4	5	6	7	8	1	2
<i>B. perfoliata</i>	⊕	⊕	⊕	●	●																									
<i>B. sphaerocarpa</i>	⊕	⊕	⊕	⊕	⊕	⊕																								
<i>B. leucantha</i>																														
<i>B. pendula</i>																														
<i>B. alba</i>																														
<i>B. macrocarpa</i>	⊕	⊕	⊕	⊕	⊕	⊕																								
<i>B. cinerea</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
<i>B. bracteata</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
<i>B. leucophylla</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
<i>B. lanceolata</i> <sup>h</sup>																														
<i>B. nuttalliana</i>																														
<i>B. australis</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
<i>B. calycosa</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
<i>B. hirsuta</i>																														
<i>B. tectoria</i>																														
<i>B. lecontei</i>																														
<i>B. arachnifera</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		
<i>B. simplicifolia</i>	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕		

<sup>a</sup> Data on flavones, flavonols and dopag are more significant in that a blank indicates a probable absence. In other groups of flavonoids the data are fragmentary. Flowers have not been analyzed in a number of species. Species are grouped into what are believed to be phyletic series.

<sup>b</sup> Flavonols: 1, luteolin; 2, apigenin; 3, luteolin 7-β-D-glucoside; 4, apigenin 7-β-D-glucoside; 5, apigenin 7-β-D-glucoside; 6, apigenin 7-glucoside; 7, unidentified compound similar to 3; 8, apigenin monoglucoside; 9, unidentified.

<sup>c</sup> Flavonols: 1, quercetin 7-β-D-glucoside (ratin); 2, rutin 7-β-D-glucoside; 3, kaempferol diglycoside; 4, quercetin 3-β-D-glucoside; 5, kaempferol 3-glucoside; 6, quercetin; 7, kaempferol; 8, quercetin 7-β-D-glucoside; 9, quercetin 3-β-D-glucoside; 10, quercetin 3,7-di-β-D-glucoside; 11, quercetin 3,7-di-β-D-glucoside; 12, quercetin 3,7-di-β-D-glucoside.

<sup>d</sup> Isoflavones: 1, sphaerobioside; 2, pseudobaptisin; 3, genistein; 4, oroboside; 5, formononetin; 6, tectoridin; 7, biochanin A; 8, other isoflavones.

<sup>e</sup> Flavonols: 1, eriodictyol; 2, narigenin.

<sup>f</sup> Dopa = 5,6-dihydroxyphenylalanine.

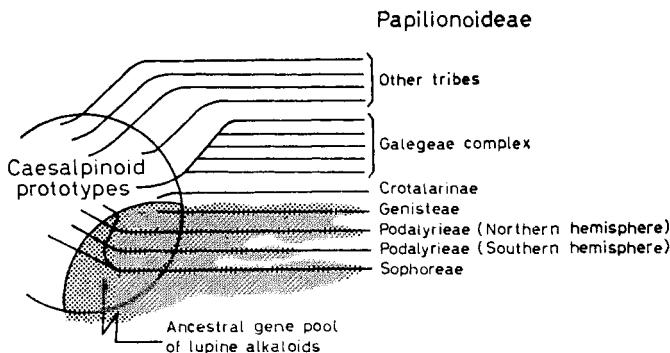
<sup>g</sup> Flowers only.

<sup>h</sup> *B. pendula* is perhaps best treated as a regional variant of *B. leucantha*; *B. bracteata* is perhaps better treated as a variety of *B. leucophylla*.

<sup>i</sup> Includes the regional variant, *B. elliptica*.

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

(e.g. sparteine). Tetracyclic lupine alkaloids are believed to be biosynthetic precursors to the tricyclic type and this information, along with the morphological and ecogeographical data discussed above, again speaks for a *Thermopsis-like* element as the older phyletic stock. That both genera are in turn related to the tribe Sophoreae may be ascertained from the fact that *Sophora* possesses, as minor alkaloids, at least five of the approximately eight alkaloids common to *Baptisia* and *Thermopsis* (Table 2). Lupine alkaloids are notably absent in the southern hemisphere Podalyrieae, only a single compound (lupanine) reported to date.



*Figure 5.* Tribes of the Leguminosae which contain lupine alkaloids, arranged according to hypothetical phyletic affinities<sup>66</sup>

I have chosen *Baptisia* and *Thermopsis* to discuss the bearing of microchemical data on systematic problems because of my familiarity with the morphology of the groups. However, the kind of phyletic problems raised by these two genera can be found again and again in vascular plants. For example, *Vicia* and *Lathyrus* are distinguished primarily by characters of the ovary. If it were not for an intuitive weighting of these two characters, certain elements of *Vicia* might easily go into *Lathyrus* and *vice versa*. That is, a purely phenetic treatment would tend to obscure the phyletic relationships. Bell's work<sup>67</sup> on the distribution of free amino acids in these two genera is particularly instructive; indeed, his data clearly reinforce the treatment of these taxa as distinct phyletic lines, implicit in the *a posteriori* judgements, from morphological data, which treat the taxa as distinct genera.

One of the more intriguing studies in which micromolecular data have provided phyletic insights, perhaps unobtainable by purely exomorphic means, is that of McClure<sup>68</sup> on the Lemnaceae (Duckweed family). A number of flavonoids occur in the four genera, *Spirodela*, *Lemna*, *Wolffia*, and *Wolffiella*, which comprise the family, and the distribution of these compounds among the genera is particularly interesting.

To appreciate the data, however, a brief taxonomic account of the family must be given. It is a wholly aquatic family belonging to the monocotyledonous groups. The included species are highly reduced, both in size and outward complexity. Some of the species (e.g., *Wolffia* spp.) measure less than 2 mm across, being devoid of roots, apparent stems or leaves; indeed, they might be described as minute, green vegetative balls (the species flower but rarely). Most workers agree that the Lemnaceae has been derived from

Table 2. Distribution of "Baptisia-type" alkaloids among legume genera†

Genus	Cytisine	Methylcytisine	Anagyrine	Sparteine	Lupanine	Baptifoline	Hydroxysparteine	Thermosine	No. of other alkaloids reported
Baptisia	++	++	++	+++	+	+	+	+	1-2
Thermopsis	+++	++	++	+++	+	+	+	+	5
Anagyris	++	++	++	++	++	++	++	++	0
Piptanthus	++	++	++	++	++	++	++	++	2
Podalyria‡	++	++	++	++	++	++	++	++	0
Genista	++	++	++	+++	++	++	++	++	2-4
Retama	++	++	++	++	++	++	++	++	4
Cytisus	++	++	++	++	++	++	++	++	0
Ulex	++	++	+	++	++	++	++	++	0
Spartium									0
Sarothamnus									1
Lupinus	+			+					9
Hovea									0
Tempietonia									0
Adenocarpus									0
Laburnum	++								0
Sophora	+	+	+	+	++	++	+	+	6
Ammodendron									1
Ammothamnus									2
Virgilia									2
Ormosia									3-4

† If the alkaloid is a major component for the genus as a whole this is indicated by ++ notation; minor alkaloids by +. (from Grammer and Turner\*).

‡ The only southern hemisphere genus of the Podalyriace known to have such alkaloids, although a number of other genera have been examined (White, 1951).

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

terrestrial ancestors and that within the family there has been a phyletic trend toward increasing simplicity through reduction<sup>69-71</sup>. Thus *Spirodela*-like species are believed to have given rise to *Lemna*-like species and the latter to the most highly reduced, *Wolffia* and *Wolfiella* (Figure 6).

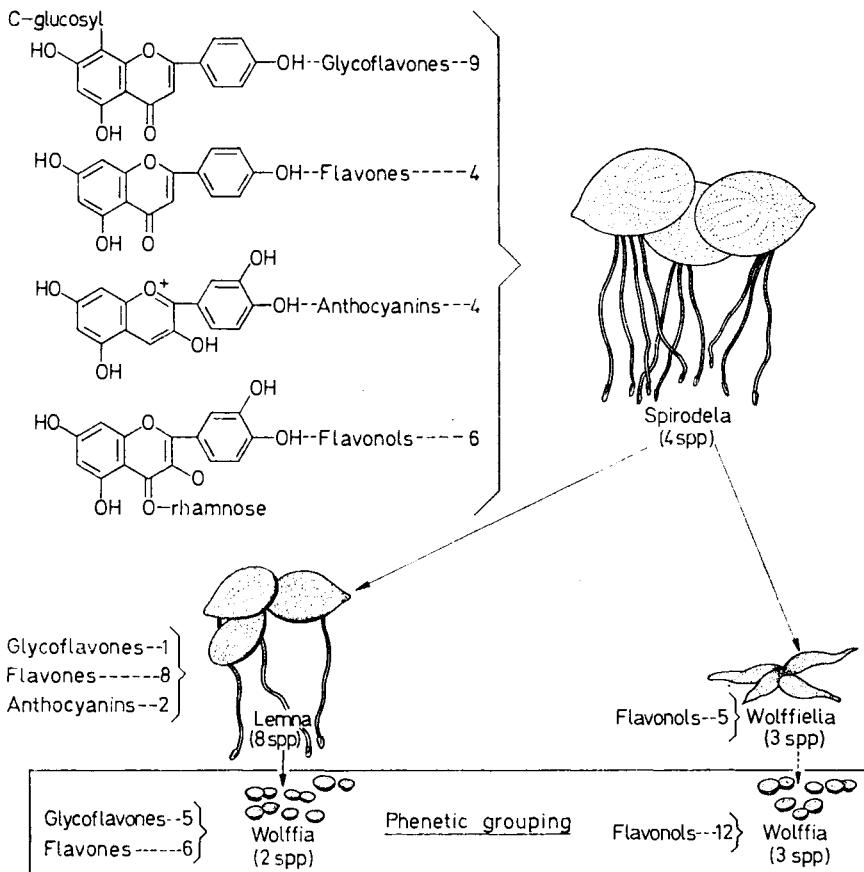


Figure 6. Hypothetical relationships in Lemnaceae. Morphological reduction is believed to have given rise to *Wolffia* and *Wolfiella* from ancestral prototypes similar to *Spirodela*. This reduction is also paralleled by a reduction in the kinds of secondary compounds produced. However, the chemical data suggest that *Wolffia*, as classically circumscribed, is biphyletic, the group of species without flavonols coming from a *Lemna*-like line; the group possessing only flavonols arising from some ancestral line perhaps unrelated to the *Lemna* group. (Based on data from McClure<sup>68</sup>)

The chemical data are intriguing because these also point to a phyletic reduction series, paralleling that of the morphological. *Spirodela* contains four flavonoid types: (i) anthocyanins, (ii) flavones, (iii) glycoflavones, and (iv) flavonols: *Lemna* possesses (i), (ii), and (iii); *Wolffia* (ii) and (iii); and *Wolfiella* only (iv) (Figure 6). The flavonoids of *Wolffia* and *Wolfiella* are particularly interesting. These two genera are distinguished primarily by their vegetative shapes, *Wolffia* being more nearly isodiametric, *Wolfiella* being more nearly linear. *Wolffia*, then, appears to be a taxonomic category

Table 3. Distribution of flavonoids in *Wolffia* (Adopted from McClure, 1964)

Species	Flavones†					Glycosflavones†					Flavonols†‡									
	1	2	3	1	2	3	4	5	1	2	3	4	5	6	7	8	9	10	11	12
<i>W. punctata</i>									+	+	+	+	+	+	+	+	+	+	+	+
<i>W. microscopica</i>									+	+	+	+	+	+	+	+	+	+	+	+
<i>W. papulifera</i>									+	+	+	+	+	+	+	+	+	+	+	+
<i>W. columbiana</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>W. arrhiza</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

† Numbers refer to different structural types of the compounds concerned; + sign denotes presence of the compound.  
‡ *Wolffia*, sometimes included in *Wolffia*, possesses only flavonoids. Flavonols are absent in *Spiriodella*, while flavones are found in both genera.

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

which has been erected to include those species of the Lemnaceae which are devoid of exomorphic characters. The flavonoid data (*Table 3*), however, suggest that this grouping is perhaps not a monophyletic one, in fact, it suggests that *Wolffia*, as presently constituted, is biphyletic, containing species derived through a *Lemna-like* line and another stemming from a *Wolffiella-like* element. It will be most interesting to see if a parallel reduction series might not be found in yet other chemical groups within this complex. Most important for the purposes of this symposium, we have in *Wolffia* an example where the chemical data have suggested a phyletic model which could hardly have been proposed by exomorphic data, but surely an exceptional case among the higher plants generally, for most chemical data are likely to be used to support or deny those phyletic models already proposed by plant systematists<sup>3</sup>.

Clearly, in the several examples cited above, the chemical data, after structural determinations have been made, must be evaluated in different ways, effective interpretation depending in large measure upon the structures concerned. Where the secondary compounds are quite different it may perhaps be inferred that the metabolic pathways leading to their production are quite different and, consequently, more phylogenetic emphasis might be placed upon distributional data of such compounds than would ordinarily be expected. An example of this type has been the phylogenetic conclusions drawn out of surveys for the red and yellow pigments (betacyanins and betaxanthins). These pigments are confined to the order Centrospermae and, unlike the common anthocyanins and carotenoids which comprise the red and yellow pigments of most other angiosperm groups, the betacyanins and betaxanthins are nitrogenous compounds which appear to be unrelated both chemically and biosynthetically to the former. Mabry<sup>72</sup> has recently reviewed the literature bearing on the phylogenetic implication of these data and pertinent comments from his paper follow:

"... The taxonomic reliability of the betacyanins and betaxanthins rests not only on their chemically unique structures and the observation that their distributions are limited to the Centrospermae, but also on the fact that the pigments are mutually exclusive with the much more widely distributed anthocyanin pigments (although other classes of flavonoid pigments occur in the Centrospermae). For these reasons, along with morphological considerations, the Order Centrospermae has been recognized as containing the ten betacyanin families: Chenopodiaceae, Portulacaceae, Amaranthaceae, Nyctaginaceae, Phytolaccaceae, Stegnospermaceae (often treated as a subfamily of the Phytolaccaceae), Aizoaceae, Cactaceae and Didieraceae. . . .

"... Betacyanins and betaxanthins apparently developed solely in the Centrospermae at a very early time, perhaps even before the anthocyanins appeared generally in the angiosperms. Several classes of flavonoids other than the anthocyanins do occur in the betacyanin families. Although detailed knowledge of the biosynthetic relationships among the flavonoids is not available, it is known, however, that most, if not all, flavonoids including the anthocyanins are formed from similar precursors, presumably chalcones. Furthermore, the *in vivo* transformations of either chalcones or other more highly oxidized flavonoids to anthocyanins is probably not enzymatically complex. Therefore it appears that the betacyanin families did not evolve pathways to the anthocyanins because the betacyanin pigments perform the same function in these families that is filled by anthocyanins in other families. . . ." "... While the Centrospermae is perhaps not any better defined by the betacyanin data, the latter do, however, contradict the *arrangement* of the families as proposed by Hutchinson, for he would disperse several of the taxa among quite different, supposedly phyletic, lines. . . ."

The betacyanin story is one, if not the best, example of how micro-molecular data might bear upon phylogeny, particularly at the family level or higher, where new data are especially needed. A similar study has been the

### B. L. TURNER

surveys of Kjaer<sup>73</sup> and others for the sulphur-containing isothiocyanates. These compounds have been found repeatedly in the Capparidaceae, Cruciferae, Moringaceae, and Resedaceae but not in the Papaveraceae, hence the data support Takhtajan's views that the latter family belongs to a different phyletic line. These data, like those of the betacyanins, are also contrary to Hutchinson's treatment, for Hutchinson would place the woody Capparidaceae and Moringaceae in a fundamentally different phyletic group from that of the predominantly herbaceous Cruciferae and Resedaceae. Unfortunately, the isothiocyanates also occur sporadically in several widely separated families such as the Euphorbiaceae, Phytolaccaceae and Plantaginaceae. Exceptions to the betacyanin story have not been found, in spite of rather wide surveys for the pigment.

Several workers have suggested phyletic interpretations on the basis of surveys of a similar type, but most such attempts have left too many sampling gaps to be meaningful, or else the compounds were too erratic in their occurrence to justify undue enthusiasm<sup>4-6,74</sup>. Distributional surveys for secondary compounds will become much more meaningful after knowledge of their biosynthetic origins has accumulated. This point has been repeatedly emphasized by a number of workers<sup>74</sup> and in at least one recent text<sup>75</sup> on phytochemistry the secondary compounds are arranged according to the metabolic pathways leading to their production; *e.g.*, acetic acid ( $C_2$ )<sub>n</sub> derivatives; isoprenoid ( $C_5$ )<sub>n</sub> derivatives; shikimic acid derivatives, *etc.*

In summary, micromolecular data of a chromatographic nature may be of considerable value in the assessment of hybrid problems, but their utility in resolving positional problems will almost certainly demand the chemical determination of the compounds concerned, along with biosynthetic information leading to their production. A similar parallelism will be noted under the discussion of Macromolecular Systematics which follows, except that the phyletic implications which evolve out of the comparative studies of structurally determined protein and polynucleotides (*i.e.*, their residue sequences) are much more profound.

## MACROMOLECULAR SYSTEMATICS

### Serology

Many plant taxonomists prior to 1950 thought of serology and chemical taxonomy as synonymous. Alston and Turner<sup>3</sup> recently reviewed the contribution of serology to systematics and concluded that, while catalytic to positive thinking in some instances, it apparently created more disinterest in "chemistry" as an approach to plant taxonomy than it did enthusiasm. Moritz<sup>76</sup>, in a sobering review of some special features of serobotanical work has called attention to the fact that serology "is a zoological method" and that certain technical difficulties in serobotany make the approach unattractive to botanists generally.

In recent years, however, the long-favoured "precipitin" test has been replaced by the more qualitative methods of gel immunoelectrophoresis and this method, among others, has been successfully applied to selected problems. For example, Klozova and Kloz<sup>77</sup> have used immunochemical methods to detect the hybrid *Phaseolus vulgaris*  $\times$  *P. coccineus*, the F<sub>1</sub> possessing

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

a complementation of the distinctive parental protein lines, much as found in the flavonoids in hybrid *Baptisia*s. (Interestingly enough, similar techniques could not distinguish between 14 species of *Baptisia*, in spite of the fact that 11 or more distinctive arcs could be detected in each of their spectra.) But, at best, the detection of interacting macromolecular bands by immunogenetic techniques reflects the activity of relatively few genes and, viewed in this light, it is doubtful that this approach, *taken alone*, will contribute significantly to problems of plant phylogeny, although it has high value for distinguishing among presumably homologous proteins<sup>78</sup>.

### Protein-band comparisons

Unlike the secondary compounds discussed above, detailed structural information on macromolecular compounds is only now becoming generally available. To date, relatively complete structural data have been compiled for only a few kinds of proteins and the complete amino-acid sequence has been determined for only eight or so<sup>79</sup>, in spite of the fact that several hundred or more have been isolated and/or detected by electrophoretic techniques<sup>80</sup>. Most important for the purposes of this review, the amino acid sequence (primary structure) is known for only one or two plant proteins, cytochrome *c*<sup>81</sup> and papain (tentative sequence only)<sup>82</sup>, consequently most of the more enlightening macromolecular approaches, at least for higher plants generally, remain in the future.

Kaplan<sup>83</sup>, however, has pointed out that sequential studies of proteins need not be prerequisite for their use in phyletic studies. Other data, at present more readily obtained, such as temperature stability, electrophoretic mobility, immunological characteristics, coenzyme-enzyme relationships, etc., of appropriate enzymes can also be used in making phyletic dispositions.

Macromolecular compounds, moreover, like the secondary compounds, even without the more sophisticated characterizations, can be used to detect hybridization. Numerous techniques have been developed for the detection of protein bands by electrophoretic methods and these have been used with considerable success, especially by zoologists. Manwell *et al.*<sup>84,85</sup> have reviewed the utility of haemoglobin bands for the detection of hybrids and more recently Crenshaw<sup>86</sup> claims to have detected introgression in turtles by this approach. In fact, these techniques, for certain protein types such as the seed globulins and selected enzymes, are easier to apply and provide data more rapidly than do the techniques of paper chromatography<sup>87</sup>. Indeed, the plant taxonomist is certain to be impressed with the amount of new characters which might be acquired by these approaches. For example, using the comparatively simple techniques of column chromatography and gel acrylamide electrophoresis, Saran and Radola<sup>88</sup> were able to detect over 50 water soluble proteins in the freshwater alga, *Chlorella pyrenoidosa*. More recently, techniques have been developed for two-dimensional gel electrophoresis of proteins and, if specific protein types can be isolated, there exist fast "fingerprinting" techniques, using tryptic digests and thin layer polyacrylamide, for the detection of protein fragments or peptides<sup>89</sup>.

## B. L. TURNER

Application of protein-band data to specific taxonomic problems in plant groups are relatively rare, although a number of workers have called attention to the systematic potential of such studies<sup>90,91</sup>. One of the more interesting contributions has been that of Johnson and Hall<sup>92</sup> where protein-band data were used to advantage in their investigation of the complex genome relationships in *Triticum*.

### DNA hybridization

Perhaps the most spectacular chemosystematic development to come out of the past decade has been the techniques developed by Bolton and McCarthy<sup>93</sup> and others in which DNA is extracted from an organism and made to "hybridize" *in vitro* (presumably through nucleotide-sequence complementations) with the extracted DNA (or RNA) of another organism. The taxonomic implications of this work have been reviewed by several workers<sup>94,95</sup> and it need only be noted here that comparative data of this sort until recently (1965), have been available only for species of the animal kingdom and microorganisms. However, Bolton<sup>96</sup> and his colleagues have now obtained such data for certain higher plants. Surprisingly, the preliminary results indicate considerable genetic divergence between what are thought to be closely related taxa (*e.g.*, *Pisum*, *Vicia*, and *Phaseolus*, all belonging to the family *Leguminosae*). Thus Bolton notes that "within the family *Leguminosae* the relative diversity as measured by these DNA interactions is at least as great as exists among orders of mammals as distantly related as men and mice". If the methods are developed further, *e.g.*, by the techniques of McLaren and Walker<sup>97,98</sup> and those of Gillespie and Spiegelman<sup>99</sup> we can hopefully anticipate a set of numerically expressed relationship-figures which, in theory, purport to show the total similarities and/or dissimilarities between whole genomes. Nucleic acid hybridization, with present techniques, will probably not be very useful for those seeking information about the relationship between and among closely related plant taxa. But the approach, if data are obtained comparable to those obtained for the animal kingdom, is perhaps the only realistic one in the foreseeable future for getting at objectively determined relationships of families and the higher categories.

### Comparative structure

It would be presumptuous to review here the background and voluminous literature which bear upon the more sophisticated implications of macromolecular systematics. This has been covered energetically by Lanni for microorganisms and viruses, and the recent symposium on *Evolving Genes and Proteins*<sup>13</sup> has brought together the views of experts in the field. These studies, while not immediately applicable to plant systematics, are nevertheless pertinent for their extrapolatory value and, consequently, the implications of this kind of work will be touched on briefly in the present paper.

Concepts arising out of molecular biology, from the standpoint of organismally-centred systematists, have had a certain familiarity, for it seems that molecularly oriented workers have chanced upon the evolutionary

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

theory quite independently of Darwin. Indeed, had the concept of organic evolution not been proposed by organismally-centred biologists, it almost certainly would have been pronounced by present-day chemists. In fact, many university courses in biochemistry are organized, often unconsciously, along an evolutionary framework, *i.e.*, simple or "primitive" metabolic pathways are discussed first, and upon these the more complex systems are developed. That this framework may be correlated with what the biologist considers to be primitive versus advanced organisms often comes as a surprise to the chemists themselves.

Contemplation of molecular evolution is a rewarding intellectual experience, for any number of parallels can be made with comparable phenomena at the organism level (or *vice versa*). Thus there ought to be a field entitled "molecular ecology" in which the "competition" between enzymes and substrates under varying cellular environments might be studied; or a subfield entitled "molecular paleontology" (see additional comments, below) in which attempts are made to "dig out" of *extant*<sup>†</sup> genomes (either directly by the comparison of nucleotide sequences, or by inference from the primary structure of proteins) those macromolecular units now covered-over with the molecular debris of the ages, *etc.*

Certainly it is difficult to consider the systematic implications of the protein work of Ingram<sup>102</sup>, Zuckerkandl and Pauling<sup>103</sup>, and Margoliash and coworkers on cytochrome *c* without a certain sense of awe. The work of Margoliash and coworkers on cytochrome *c* is perhaps typical and this will be briefly reviewed here.

Cytochrome *c*, a component of the terminal respiratory chain of enzymes in aerobic organisms, is widely distributed in both the animal and plant kingdoms. The enzyme has been obtained in pure form from a number of species and comparative study of the primary structures has shown that closely related taxa possess similar cytochrome *c* enzymes. By inference from these data it appears certain that:

1. cytochrome *c* enzymes in all organisms are homologous;
2. by extrapolation, they arise from homologous gene loci;
3. each possesses an evolutionary history;
4. they have diverged from some ancestral prototype;
5. "phylogenies" of proteins can be traced;
6. inferred primitive type molecules can be reconstructed (both on paper and in the test tube);
7. molecular phylogenies can be tested statistically and possibly experimentally<sup>105</sup>.

In addition to the comparative studies on cytochrome *c*, extensive information exists on the amino-acid sequences of the haemoglobin molecule. Zuckerkandl and Pauling<sup>103</sup> have, among others, recently reviewed the systematic import of these data and it appears that haemoglobin has evolved at a rate of about 1 amino-acid change or mutation every 10 million years.

<sup>†</sup> Extinct organisms may also be analysed for their fossil macromolecules<sup>100</sup>, *e.g.*, Foucart *et al.*<sup>101</sup> have reported on the fossil proteins of graptolites from paleozoic time, concluding, by amino acid analysis, that the major macromolecular units were scleroproteic in nature. The expected residues for chitin (as it is known today) were not found, nor was there any trace of cellulose.

#### B. L. TURNER

A "unit evolutionary period" of 22·6 million years has been estimated for the mutation and/or fixation of an amino-acid in cytochrome *c*<sup>104</sup>. This figure was based on a paleontologically determined period of 280 million years (*i.e.*, the estimated divergence of the avian and mammalian lines). With this time unit as a standard, it is reckoned that the divergence of the fish from the bird-mammal line occurred about 460 million years ago; vertebrates from invertebrates at about 680 million years; and fungi from this animal-line about 1000 million years. As indicated, such time-units and extrapolations from these, have been established by reference to what is known of the fossil record. It will be most interesting to see what comes out of similar studies in the flowering plants, for a fossil record pointing to clear cut evolutionary lines in this group is lacking.

Comparative studies of proteins are in their infancy and it is unlikely that any great accumulation of data which bear upon plant systematics generally will be had in this generation. However, biochemical techniques are developing rapidly and even if the instruments and know-how are not available to systematists of this day, it is likely that they will become so tomorrow. Relatively rapid methods have already been developed for the analysis of peptides derived from tryptic digests of protein<sup>106</sup> and as the techniques<sup>107</sup> and instrumentation for peptide analysis become more sensitive, the high speed computer is certain to resolve the more tedious problems of interpretation and/or protein reconstruction<sup>108</sup>.

That future developments in chemosystematics might indeed be part of a Brave New World may be ascertained from a rather remarkable contribution by Pauling and Zuckerkandl<sup>105</sup>. In an imaginative article entitled "Chemical Paleogenetics: Molecular 'Restoration Studies' of Extinct Forms of Life" these authors comment:

"... Paleobiochemistry, through molecular restoration studies on the bases of existing related polypeptide chains provides the means of investigating the structure of such components for any part of the genome of extinct organisms. [for] ... once the structures of ancestral polypeptide chains are known it will in the future be possible to synthesize these presumed components of extinct organisms ... [and] as information about various paleogenes belonging to a given group of extinct organisms will accumulate, some deductions concerning these organisms will be possible in relation to levels of biological integration higher than the level of individual macromolecules. When a fossil record is available, knowledge about the organisms concerned will go far beyond what has so far been believed possible. Important information will also be provided about forms that have left no fossil record whatsoever, such as many soft-bodied animals..."

#### EPILOGUE

There is certain to come a day when the present gap between organismally centred and molecularly centred systematics become sufficiently interfused so as to form a conceptual framework from which will grow remarkable insights into the whole of biological evolution. Neither group will develop these concepts alone, for the field of systematics is three-dimensional; there is no single focus. The molecular approach has done something for systematics unanticipated by taxonomists themselves; it has made systematics the ultimate approach of nearly all biological work. The molecular questions, what is it?, where does it occur? and what is it doing?, are taxonomic questions. The higher questions, when?, where? and how did it come about?

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

are phylogenetic questions. Taken together they constitute the spirit, if not the essence, of chemosystematics; indeed, all of biology.

*Parts of this paper were presented at a symposium on "Newer Trends in Taxonomy" sponsored by the National Institute of Sciences of India, New Delhi, 1965. Chemosystematic studies on Baptisia were supported by National Institute of Health Grant—11111. I am grateful to R. E. Alston, J. Harborne, and T. J. Mabry for helpful suggestions in the preparation of this presentation.*

### References

- <sup>1</sup> L. Constance. *Taxon* **13**, 257 (1964).
- <sup>2</sup> B. L. Turner. In *Newer Trends in Taxonomy*. National Institute of Sciences, New Delhi, India (1966).
- <sup>3</sup> R. E. Alston, and B. L. Turner. *Biochemical Systematics*. Prentice Hall, Inc., Ingleside, (1963).
- <sup>4</sup> R. Hegnauer. *Chemotaxonomie der Pflanzen*. Vol. 1-5, Birkhauser, Verlag (1962-66).
- <sup>5</sup> C. A. Leone. (Ed.) *Taxonomic Biochemistry and Serology*. Ronald Press Co., New York (1964).
- <sup>6</sup> T. Swain. (Ed.) *Chemical Plant Taxonomy*. Academic Press, London (1963).
- <sup>7</sup> R. E. Alston, T. J. Mabry, and B. L. Turner. *Science, N.Y.*, **142**, 545 (1963).
- <sup>8</sup> H. Erdtman. In *Chemical Plant Taxonomy*. Academic Press, London (1963).
- <sup>9</sup> S. von Bortitz. *Biol. Zbl.* **83**, 752 (1964).
- <sup>10</sup> B. H. Hoyer, B. J. McCarthy, and E. T. Bolton. *Science, N.Y.*, **144**, 959 (1964).
- <sup>11</sup> F. Lanni. In *Plant Virology*. University of Florida Press (1964).
- <sup>12</sup> K. N. Saxena, B. R. Seshachar, and J. R. Gandhi. *Syst. Zool.* **14**, 33 (1965).
- <sup>13</sup> V. Bryson, and H. J. Vogel. (Eds.) *Evolving Genes and Proteins*. Academic Press, London (1965).
- <sup>14</sup> R. E. Alston. *Lloydia*, **28**, 300 (1965).
- <sup>15</sup> R. R. Sokal and P. H. A. Sneath. *Principles of Numerical Taxonomy*. W. H. Freeman and Co. (1963).
- <sup>16</sup> R. E. Alston. *Taxon*, **14**, 84 (1965).
- <sup>17</sup> R. D. Gibbs. In *Chemical Plant Taxonomy*. Academic Press, London (1963).
- <sup>18</sup> H. Erdtman. Personal Communication.
- <sup>19</sup> S. Rangaswami. In Symposium On Newer Trends in Taxonomy (Abstracts). Nat. Inst. Sci. India, New Delhi (1966).
- <sup>20</sup> M. J. Sackin, and P. H. A. Sneath. *Biochem. J.* **96**, 70 (1965).
- <sup>21</sup> P. H. Davis, and V. H. Heywood. *Principles of Angiosperm Taxonomy*. Oliver and Boyd, London (1963).
- <sup>22</sup> R. R. Sokal, J. H. Camin, F. J. Rohlf, and P. H. A. Sneath. *Syst. Zool.* **14**, 237 (1965).
- <sup>23</sup> E. Mayr. *Syst. Zool.* **14**, 73 (1965).
- <sup>24</sup> W. Hennig. *A. Rev. Ent.* **10**, 97 (1965).
- <sup>25</sup> V. H. Heywood, and J. McNeill. (Eds.) *Phenetic and Phyletic Classification*. E. W. Classey, Ltd. Hampton (1964).
- <sup>26</sup> J. H. Camin, and R. R. Sokal. *Evolution* **19**, 311 (1965).
- <sup>27</sup> E. Mayr. *Animal Species and Evolution*. Harvard University Press, Cambridge (1963).
- <sup>28</sup> E. A. Bell, and L. Fowden. In *Taxonomic Biochemistry and Serology*. Ronald Press Co., New York (1964).
- <sup>29</sup> N. T. Mirov. *Lloydia* **26**, 117 (1963).
- <sup>30</sup> J. W. McClure, and R. E. Alston. *Nature, Lond.* **201**, 311 (1964).
- <sup>31</sup> E. Anderson. *Introgressive Hybridization*. John Wiley & Sons, New York (1949).
- <sup>32</sup> R. E. Alston, and B. L. Turner. *Proc. natn. Acad. Sci. U.S.A.* **48**, 130 (1962).
- <sup>33</sup> R. E. Alston, and B. L. Turner. *Am. J. Bot.* **50**, 159 (1963).
- <sup>34</sup> J. Clausen, and W. M. Hiesey. *Proc. natn. Acad. Sci. U.S.A.* **46**, 484 (1960).
- <sup>35</sup> H. H. Smith, and D. V. Abashian. *Am. J. Bot.* **50**, 135 (1963).
- <sup>36</sup> E. Nowacki, and D. B. Dunn. *Genet. pol.* **5**, 47 (1964).
- <sup>37</sup> H. H. Smith. *Am. Nat.* **99**, 73 (1965).
- <sup>38</sup> C. R. Parks. *Am. J. Bot.* **52**, 309 (1965).

B. L. TURNER

- <sup>39</sup> J. McHale, and R. E. Alston. *Evolution*. **18**, 304 (1964).
- <sup>40</sup> D. M. Smith, and D. A. Levin. *Am. J. Bot.* **50**, 952 (1963).
- <sup>41</sup> G. L. Stebbins, B. L. Harney, E. L. Cox, J. N. Rutger, G. Jelencovic, and E. Yagil. *Am. J. Bot.* **50**, 830 (1963).
- <sup>42</sup> A. O. Taylor. *New Phytol.* **63**, 135 (1964).
- <sup>43</sup> A. M. Torres, and D. A. Levin. *Am. J. Bot.* **51**, 639 (1964).
- <sup>44</sup> P. M. Harney, and W. F. Grant. *Am. J. Bot.* **51**, 621 (1964).
- <sup>45</sup> B. G. Brehm, and M. Ownbey. *Am. J. Bot.* **52**, 811 (1965).
- <sup>46</sup> D. A. Levin. *Am. J. Bot.* **53**, 238 (1966).
- <sup>47</sup> G. Hunter. Unpublished work, Univ. of Texas. (1965).
- <sup>48</sup> O. Stromnaes, and E. D. Garber. *Bot. Gaz.* **105**, 363 (1963).
- <sup>49</sup> N. T. Mirov. In *The Physiology of Forest Trees*. Ronald Press Co., New York (1958).
- <sup>50</sup> N. T. Mirov. *Techn. Bull. U.S. Dep. agric.* No. 1239. 1 (1961).
- <sup>51</sup> M. H. Bannister, H. V. Brewerton, and I. R. C. MacDonald. *Svensk Papp-Tidn.* **62**, 567 (1959).
- <sup>52</sup> M. B. Forde. *N. Z. J. Bot.* **2**, 53 (1964).
- <sup>53</sup> D. E. Stone, G. A. Adrouny, and S. Adrouny. *Brittonia* **17**, 97 (1965).
- <sup>54</sup> R. E. Alston. In *Biochemistry of Phenolic Compounds*. Academic Press, London (1964).
- <sup>55</sup> D. B. Horne. Doctoral Dissertation, University of Texas (1965).
- <sup>56</sup> K. Baetcke. Doctoral Dissertation, University of Texas (1965).
- <sup>57</sup> E. von Rudloff. *Gas Chromat.* **3**, 390 (1965).
- <sup>58</sup> F. M. Couchman, and E. von Rudloff. *Can. J. Chem.* **43**, 1017 (1965).
- <sup>59</sup> D. B. Horne. *Am. J. Bot.* **51**, 687 (1964).
- <sup>60</sup> R. E. Alston, H. Rösler, K. Naifeh, and T. J. Mabry. *Proc. natl. Acad. Sci. U.S.A.* **54**, 1458 (1965).
- <sup>61</sup> W. L. Ellison, R. E. Alston, and B. L. Turner. *Am. J. Bot.* **49**, 599 (1962).
- <sup>62</sup> H. Lorenz, and J. Scholz-Schaeffer. *Z. Pflüücht.* **52**, 13 (1964).
- <sup>63</sup> H. R. Holbo, and H. N. Mozingo. *Am. J. Bot.* **52**, 970 (1965).
- <sup>64</sup> R. E. Alston, et al. Unpublished work;  
J. Harborne. Unpublished work.
- <sup>65</sup> M. F. Cranmer, and T. J. Mabry. *Phytochemistry* **5**, 1133 (1966).
- <sup>66</sup> M. F. Cranmer, and B. L. Turner. *Evolution* **20**, (1967). In press.
- <sup>67</sup> E. A. Bell. In *Comparative Phytochemistry*. Academic Press, London (1966).
- <sup>68</sup> J. McClure. Doctoral Dissertation, The University of Texas (1964).
- <sup>69</sup> A. Arbor. *Proc. R. Soc., Ser. B* **91**, 96 (1920).
- <sup>70</sup> A. Lawalree. *Bull. Soc. R. Bot. Belg.* **77**, 27 (1944).
- <sup>71</sup> A. L. Takhtajan. *Essays on the Evolutionary Morphology of Plants*. Leningrad Univ. Leningrad, p. 62 (1954). (Trans. Multiprint, Inc., New York).
- <sup>72</sup> T. J. Mabry. In *Comparative Phytochemistry*. Academic Press, London (1966).
- <sup>73</sup> A. Kjaer. In *Comparative Phytochemistry*. Academic Press, London (1966).
- <sup>74</sup> T. Swain. (Ed.) In *Comparative Phytochemistry*. Academic Press, London (1966).
- <sup>75</sup> C. Mentzer, and O. Fatainoff. *Actualités de Phytochimie Fondamentale*. I. Masson et Cie., Paris (1964).
- <sup>76</sup> O. Moritz. In *Taxonomic Biochemistry and Serology*. Ronald Press Co., New York (1964).
- <sup>77</sup> E. Klozova, and J. Kloz. *Biologia Pl.* **6**, 240 (1964).
- <sup>78</sup> A. C. Wilson, and N. O. Kaplan. In *Taxonomic Biochemistry and Serology*. Ronald Press Co., New York (1964).
- <sup>79</sup> G. H. Dixon. In *Essays in Biochemistry*. Vol. II. Academic Press, London (1966).
- <sup>80</sup> M. Dixon, and E. C. Webb. *Enzymes*. Longmans, Green and Co., London (1964).
- <sup>81</sup> E. Margoliash, and E. L. Smith. In *Evolving Genes and Proteins*. Academic Press, New York (1965).
- <sup>82</sup> A. Light, R. Frater, J. R. Kimmel, and E. L. Smith. *Proc. natl. Acad. Sci. U.S.A.* **52**, 1276 (1964).
- <sup>83</sup> N. A. Kaplan. In *Evolving Genes and Proteins*. Academic Press, New York (1965).
- <sup>84</sup> C. Manwell, and C. M. A. Baker. *Comp. Biochem. Physiol.* **10**, 39 (1963).
- <sup>85</sup> C. Manwell, C. M. A. Baker, and W. Childers. *Comp. Biochem. Physiol.* **10**, 103 (1963).
- <sup>86</sup> J. W. Crenshaw. *Evolution* **19**, 1 (1965).
- <sup>87</sup> D. Boulter, D. A. Thurman, and B. L. Turner. *Taxon* **15**, 135 (1966).
- <sup>88</sup> I. Sarang, and B. J. Radola. *Mh. Chem.* **96**, 1413 (1965).
- <sup>89</sup> H. Stegemann, and B. Lerch. *Analyt. Biochem.* **9**, 417 (1964).

## PLANT CHEMOSYSTEMATICS AND PHYLOGENY

- <sup>90</sup> D. J. Fox, D. A. Thurman, and D. Boulter. *Phytochemistry* **3**, 417 (1964).
- <sup>91</sup> O. Hall, and B. L. Johnson. *Hereditas* **48**, 530 (1962).
- <sup>92</sup> B. L. Johnson, and O. Hall. *Am. J. Bot.* **52**, 506 (1965).
- <sup>93</sup> E. T. Bolton, and B. J. McCarthy. *Proc. natn. Acad. Sci. U.S.A.* **48**, 1390 (1962).
- <sup>94</sup> B. H. Hoyer, B. J. McCarthy, and E. T. Bolton. *Science* **144**, 959 (1964).
- <sup>95</sup> B. H. Hoyer, E. T. Bolton, B. J. McCarthy, and R. B. Roberts. In *Evolving Genes and Proteins*. Academic Press, New York (1965).
- <sup>96</sup> E. T. Bolton, R. J. Britten, D. B. Cowie, R. B. Roberts, P. Szafranski, and M. J. Waring. In *Carnegie Instn. Yb.* **64**, 313 (1965).
- <sup>97</sup> Anne McLaren, and P. M. B. Walker. *Genet. Res.* **6**, 230 (1965).
- <sup>98</sup> P. M. B. Walker, and Anne McLaren. *J. Mol. Biol.* **12**, 394 (1965).
- <sup>99</sup> D. Gillespie, and S. Spiegelman. *J. Mol. Biol.* **12**, 829 (1965).
- <sup>100</sup> M. Blumer. *Science* **149**, 722 (1965).
- <sup>101</sup> M. F. Foucart, S. Bricteux-Gregoire, ch. Jeuniaux and M. Florkin. *Life Sci.* **4**, 467 (1965).
- <sup>102</sup> V. M. Ingram. *The Hemoglobins in Genetics and Evolution*. Columbia Univ. Press, New York (1963).
- <sup>103</sup> E. Zuckerkandl, and L. Pauling. In *Evolving Genes and Proteins*. Academic Press, New York (1965).
- <sup>104</sup> E. Margoliash, and E. L. Smith. In *Evolving Genes and Proteins*. Academic Press, New York (1965).
- <sup>105</sup> L. Pauling, and E. Zuckerkandl. *Acta chem. scand.* **17** (Suppl. No. 1), 9 (1963).
- <sup>106</sup> C. H. W. Hirst. *Ann. Rev. Biochem.* **33**, 597 (1964).
- <sup>107</sup> R. B. Merrifield. *Science, N.Y.* **150**, 178 (1965).
- <sup>108</sup> D. F. Bradley, C. R. Merril, and M. B. Shapiro. *Biopolymers* **2**, 415 (1964).