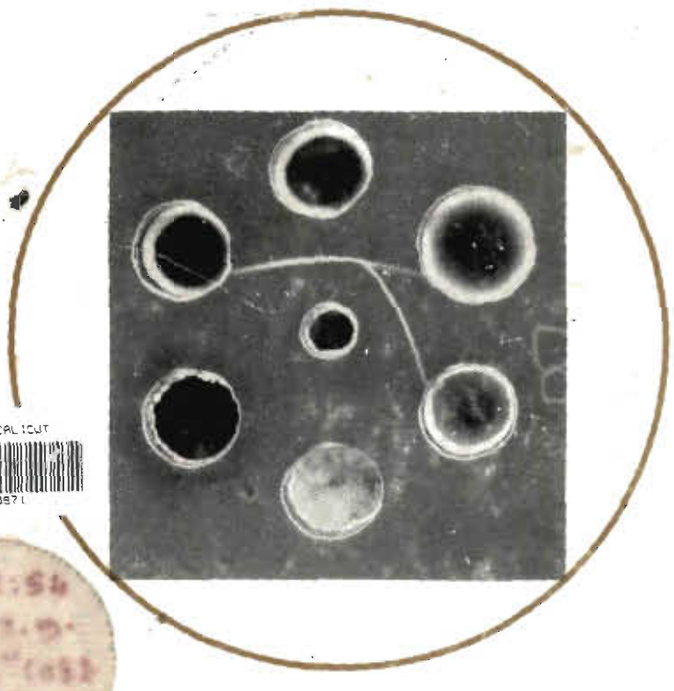
THE SYSTEMATICS ASSOCIATION SPECIAL VOLUME No. 2

Chemotaxonomy and Serotaxonomy

Edited by J. G. HAWKES



Pu[illegible]or the SYSTEMATICS ASSOCIATION by
ACADEMIC PRESS · LONDON AND NEW YORK

CHEMOTAXONOMY
AND
SEROTAXONOMY

THE SYSTEMATICS ASSOCIATION PUBLICATIONS

1. BIBLIOGRAPHY OF KEY WORKS FOR THE IDENTIFICATION OF THE BRITISH FAUNA AND FLORA
3rd edition (1967)
Edited by G. J. Kerrich, R. D. Meikle *and* Norman Tebble

2. THE SPECIES CONCEPT IN PALAEONTOLOGY (1965)
Edited by P. C. Sylvester-Bradley, B.Sc., F.G.S.

3. FUNCTION AND TAXONOMIC IMPORTANCE (1959)
Edited by A. J. Cain, M.A., D.Phil., F.L.S.

4. TAXONOMY AND GEOGRAPHY (1962)
Edited by David Nichols, M.A., D.Phil.

5. SPECIATION IN THE SEA (1963)
Edited by J. P. Harding *and* Norman Tebble

6. PHENETIC AND PHYLOGENETIC CLASSIFICATION (1964)
Edited by V. H. Heywood, Ph.D., D.Sc. *and* J. McNeill, B.Sc., Ph.D.

7. ASPECTS OF TETHYAN BIOGEOGRAPHY (1967)
Edited by C. G. Adams *and* D. V. Ager

LONDON. Published by the Association

Price per volume: *No. 1, 25s (cloth), 20s (paper); Nos. 2–6, 15s; No. 7, 30s.*
(U.S.A. No. 1, *$4.00 (cloth), $3.20 (paper); Nos. 2–6, $2.50; No. 7, $5.00.*
Free of charges)

Available from E. W. Classey Ltd., 353 Hanworth Rd., Hampton, Middlesex

THE SYSTEMATICS ASSOCIATION
SPECIAL VOLUME No. 2

CHEMOTAXONOMY AND SEROTAXONOMY

Proceedings of a Symposium held at the Botany Department, Birmingham University 15–16 September, 1967

Edited by

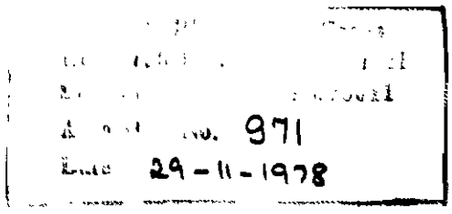
J. G. HAWKES
Department of Botany, University of Birmingham, England

1968

Published for the
SYSTEMATICS ASSOCIATION
by
ACADEMIC PRESS • LONDON • NEW YORK

ACADEMIC PRESS INC. (LONDON) LTD
Berkeley Square House
Berkeley Square
London, W.1

U.S. Edition published by
ACADEMIC PRESS INC.
111 Fifth Avenue
New York, New York 10003



Library of Congress Catalogue Card Number: 68–56050

Made and printed in Great Britain by
William Clowes and Sons, Limited, London and Beccles

Contributors

R. P. AMBLER, *Department of Molecular Biology, University of Edinburgh, Scotland.*

E. C. BATE-SMITH, *Agricultural Research Council, Institute of Animal Physiology, Babraham, Cambridgeshire, England.*

BIRGER BLOMBÄCK, *Department of Blood Coagulation Research, Karolinska Institutet, Stockholm, Sweden.*

MARGARETA BLOMBÄCK, *Department of Blood Coagulation Research, Karolinska Institutet, Stockholm, Sweden.*

D. BOULTER, *Department of Botany, University of Durham, England.*

A. J. CAIN, *Department of Zoology, University of Manchester, England.*

S. T. COWAN, *Formerly of the Public Health Laboratory Service, Central Public Health Laboratory, London, England.*

HOLGER ERDTMAN, *Department of Organic Chemistry, Royal Institute of Technology, Stockholm, Sweden.*

P. G. H. GELL, *Department of Experimental Pathology, University of Birmingham, England.*

H. J. GOULD, *M.R.C. Biophysics Research Unit, Department of Biophysics, King's College, University of London, England.*

J. B. HARBORNE, *Phytochemical Unit, The Hartley Botanical Laboratories, University of Liverpool, England.*

G. A. D. HASLEWOOD, *Biochemistry and Chemistry Department, Guy's Hospital Medical School, London, England.*

J. G. HAWKES, *Department of Botany, University of Birmingham, England.*

J. HESLOP-HARRISON, *Institute of Plant Development, University of Wisconsin, Madison, Wisconsin, U.S.A.*

R. G. HUNTSMAN, *Department of Pathology, Lambeth Hospital, London, England.*

UWE JENSEN, *Institut für Pharmakognosie, University of Kiel, West Germany.*

JOSEF KLOZ, *Institute of Experimental Botany, Czechoslovak Academy of Sciences, Prague, Czechoslovakia.*

EVA KLOZOVA, *Institute of Experimental Botany, Czechoslovak Academy of Sciences, Prague, Czechoslovakia.*

D. E. KOHNE, *Department of Terrestrial Magnetism, Carnegie Institution, Washington, D.C., U.S.A.*

J. R. NORRIS, *"Shell" Research Limited, Milstead Laboratory of Chemical Enzymology, Sittingbourne, Kent, England.*

T. F. PREECE, *Agricultural Botany Division, School of Agricultural Sciences, University of Leeds, England.*

IVOR SMITH, *Courtauld Institute of Biochemistry, The Middlesex Hospital Medical School, London, England.*

R. Y. STANIER, *Department of Bacteriology and Immunology, University of California, Berkeley, California, U.S.A.*

D. A. THURMAN, *The Hartley Botanical Laboratories, University of Liverpool, England.*

W. G. TUCKER, *Department of Botany, University of Birmingham, England.*

J. G. VAUGHAN, *Department of Biology, Queen Elizabeth College, University of London, England.*

ORGANIZING AND EDITORIAL COMMITTEE

Preface

The present volume contains papers read at the Symposium on Chemotaxonomy and Serotaxonomy held at Birmingham University on 15 and 16 September, 1967. It contains the full texts or extended summaries of the papers presented at the meeting. Sixteen* of these were read by invitation whilst six were accepted at the request of the authors. These latter papers were presented in a shortened form in the last session but have been distributed in the book in positions consistent with their subject-matter. A general discussion at the end of the Symposium was preceded by a summing-up by the Chairman for that session, and the text for this is also included.

Although the Editor realizes that it would have been valuable to present a verbatim report or summary of the discussions which took place at the Symposium, it was felt that the delay resulting from the transcription and collation of these discussions would not have been justified, in view of the urgent need for rapid publication in this quickly developing field.

In planning the Symposium, the committee divided the subject matter into protein and non-protein studies and allocated one session to serological and immunological work. The first session was devoted to papers on general protein investigations, electrophoresis, etc.; the second dealt with serology and immunology, and the third comprised papers on nucleic acids. Session 4 was mainly concerned with biochemical approaches not connected with protein studies, whilst in session 5 the speakers were asked to assess the value of biochemical and serological data in taxonomy. At all sessions but the third, papers were invited on animals, plants and micro-organisms so as to cover as wide a field as possible of biological interests. The third session, on nucleic

* A paper read by Professor O. Westphal on the "Chemical, serological and genetic classification of the Enterobacteriaceae" has had to be omitted since the manuscript was not received at the time of going to press.

acids, included papers of very general interest to biologists, in which work on a wide variety of organisms was discussed.

Whilst the original intention of the organizing committee was to present serological and biochemical techniques to the taxonomist and to show him the rapidly developing aspects of chemical taxonomy as well as the importance of chemical and serological criteria in modern systematics, the general theme of some of the contributions veered towards comparative biochemistry, rather than to taxonomy. However, the speakers in the last session very ably discussed the value of these techniques in a wider context and examined the extent to which the data could be used to throw light on phylogeny and evolutionary relationships.

The Symposium showed clearly the growing importance of biochemical and immunological approaches in modern taxonomic work. It also underlined the need for a greater understanding of biochemistry on the part of the taxonomist and a greater taxonomic understanding amongst biochemists and others in related fields who are carrying out comparative studies on a range of different organisms.

The interest shown at the Symposium was more than sufficient to justify the views of the committee and the Systematics Association that the proceedings of the meeting should be published. Since it was thought that the theme exteuded beyond the usual interests of members of the Association and would at the same time be of interest to biologists who were not members and who did not normally see the Association's publications it was decided that the book should be published by a commercial publishing house, rather than by the Association itself.

The Editor would like to record here his very grateful thanks to members of the committee who helped to organize the Symposium and who subsequently kindly agreed to serve on the editorial committee. Thanks are also due to the past President of the Systematics Association, Professor P. Sylvester-Bradley, as well as to the botanical secretary, Dr. J. McNeill, and members of the publishing committee for their help and advice in steering this book through the press.

J. G. HAWKES

Botany Department,
University of Birmingham.
June 1968

Contents

Section III: Large Molecules (Nucleic Acids)

Section IV: Other Biochemical Approaches

Section V: The Assessment of Biochemical Techniques in Taxonomy

Foreword

P. H. A. Sneath
University of Leicester

It is over thirty years since the Systematics Association was founded, and in that time it has sponsored a wide range of symposia on topics of interest to systematists. It has interpreted systematics in the widest sense and has always endeavoured to combine botanical, zoological and palaeontological aspects wherever possible. The Association is not merely a society of systematists, but is also a forum for new ideas from other fields. It has a responsibility to explore these, and to find out how discoveries in other disciplines can be applied to systematic biology. This volume, combining many applications, from the molecular to the evolutionary, is therefore directly in its tradition.

This volume contains the proceedings of a Symposium on Chemotaxonomy and Serotaxonomy held on 15 and 16 September, 1967, at the University of Birmingham. These subjects have had a long history (serotaxonomy dates back to Nuttall's work at the beginning of this century, and chemotaxonomy in scattered applications was used even earlier): but it is only in recent years that they have come to command the interest due to them, in large part because of advances in technique. One section of chemotaxonomy has arisen in the more recent past as an offshoot of molecular biology. The information in large molecules—proteins and nucleic acids—is vast and fascinating. We are beginning, also, to understand the relations between the fine genetic structure and serology and chemistry of other kinds. In all these fields information is now growing swiftly.

How will all this new knowledge be integrated into the body of systematics? And how does it relate to the scope of the systematics of the future? Description and analysis must precede synthesis, and today we are still in the first stage. We can already envisage, however, some outlines of the second stage. This is perhaps clearest in protein sequence studies. Here we can see how a coherent evolutionary theory and structure at the molecular level may be built, to compare with that

based on more familiar data. We can expect advances in chemistry as applied to fossil remains to complement it. Reading molecular detail by electron microscopy is not beyond the bounds of possibility, and this would open up enormous areas to explore. There are similar trends in the other fields covered in this volume, such as elegant chemical and serological investigations into the wild progenitors of cultivated plants. Indeed this new knowledge promises to redress the shortage of critical information which has made it so difficult in the past to make consistent and logical deductions about evolution.

Integration of this information into systematics will call for far-sighted taxonomists who are prepared to take it into their scheme of things. Much of it may be scattered in unfamiliar publications. Considerable effort will then be needed to gather it together. When the data become voluminous, computers will assist us, while biometric methods will be increasingly needed to evaluate the significance of the new discoveries. The future will be exciting, and the reader can capture some of the excitement in the essays that follow. It gives me great pleasure as President of the Association to introduce this volume, to join the growing series of the Association's publications.

SECTION I

LARGE MOLECULES (PROTEINS)

1. Primary Structure of Animal Proteins as a Guide in Taxonomic Studies

BIRGER BLOMBÄCK and MARGARETA BLOMBÄCK

Department of Blood Coagulation Research, Karolinska Institutet, Stockholm, Sweden

ABSTRACT

Studies on the primary structure of fibrinogen from different mammalian species have provided valuable information on mammalian classification. The amino acid sequences of the fibrinopeptides split off from the parent fibrinogen molecule by the action of thrombin were studied in 34 different species, mainly mammalian. Fibrinopeptides seem to be specially suitable for taxonomic studies as they constitute genetically equivalent parts of the fibrinogen from the different species and contain 17 amino acids, which can be considered as the elementary building-stones of the molecules.

The classification of species according to the amino acid sequences of the peptides agrees in general with the accepted taxonomy based on morphological data.

In the past 10–20 years we have witnessed an extremely fast development in protein chemistry. The elucidation of a protein structure which was just a dream 20 years ago can now, though with considerable effort, be made in relatively short time. As all data so far indicate that the conformation of a protein is basically determined by its primary structure, i.e. the amino acid sequence of the protein, the elucidation of protein primary structure is of utmost importance for the understanding of the relation between structure and function of proteins. Such knowledge will lead to a finer appreciation of pathological processes in which abnormal proteins are involved. Furthermore, according to current views, the primary structure of proteins is eventually determined by specific nucleotide sequences in the DNA of the structural genes. As it

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 3–20.

would be expected that evolutionary events in the living material would be accompanied by changes in the genetic message, it is, therefore, likely that comparative studies on primary structures of proteins might give information with regard to taxonomy and phylogenetic relationship, not only between molecular species but also between whole organisms.

We have the feeling that the hypothesis of DNA as the source of genetic information for protein primary structure is by many considered as the final truth. So far, however, no synthesis in a cell-free system of a specific protein primed by synthetic DNA-molecules has been incontrovertibly proven.

Whatever the relation between protein structure and genetic message, close or remote from the gene, the primary structure of a protein is a sensitive instrument for the demonstration of differences between species. Comparative studies on cytochrome C, fibrinogen and haemoglobin, for example, have already shown that this approach has great potentialities (Margoliash *et al.*, 1963; Margoliash, 1963; Doolittle and Blombäck, 1964; Blombäck *et al.*, 1965; Blombäck *et al.*, 1966; Pauling and Zuckerkandl, 1963; Hill *et al.*, 1963; Hill and Buettner-Janusch, 1965; Florkin, 1966).

We shall here summarize the comparative studies we have made on the primary structure of fibrinogen at the Karolinska Institutet in Stockholm. Before we discuss the results of these investigations we shall briefly consider the structure of the fibrinogen molecule and its transformation to fibrin (see Blombäck, 1967). A synthesis of all available data indicates that the fibrinogen molecule is built up of three polypeptide chains: A, B and C.* The molecular weight of 340 000 as determined for the protein is for a dimer, and the monomers are aligned in an end-to-end or similar fashion as schematically shown in Fig. 1. From the hydrodynamic measurements of Haschemeyer (1963), it seems likely that the A chains of the molecule have their N-termini at opposite ends, equidistant from the centre, but on the same side of the electric symmetry axis of the molecule. We have very little information with regard to the topography of other parts of the chains within fibrinogen.

When thrombin, a proteolytic enzyme from blood, acts on fibrinogen, two so-called fibrinopeptides A and B are split off from the N-terminal ends of the A and B chains respectively.

Current data indicate that splitting off of peptide A is necessary to initiate fibrin formation. Splitting off of peptide B is slower and seems

* The nomenclature α(A), β(B) and γ for the three chains has recently been recommended by the International Committee on Haemostasis and Thrombosis (Washington, D.C., Nov. 1967). The γ chain is equivalent to the chain denoted as C previously.

to be a secondary process which perhaps is of little importance for the immediate fibrin formation. The fibrinogen molecules that have been altered by proteolysis are generally known as "fibrin monomers". The fibrin monomers are extremely reactive units, which aggregate or polymerize into the structure we call fibrin (Fig. 1).

In its proteolytic action on fibrinogen thrombin has a narrow specificity of action. Only a few bonds in the protein are rapidly hydrolyzed. These are the arginyl-glycine bonds linking the fibrinopeptides A and B in fibrinogen to the rest of the fibrinogen molecule. These bonds are indicated by arrows in the schematic figure of the fibrinogen molecule shown in Fig. 2. We are of the opinion that the cleavage site is a fine reference point when comparing fibrinogen from different species. This view is based on the following facts: (1) Thrombin

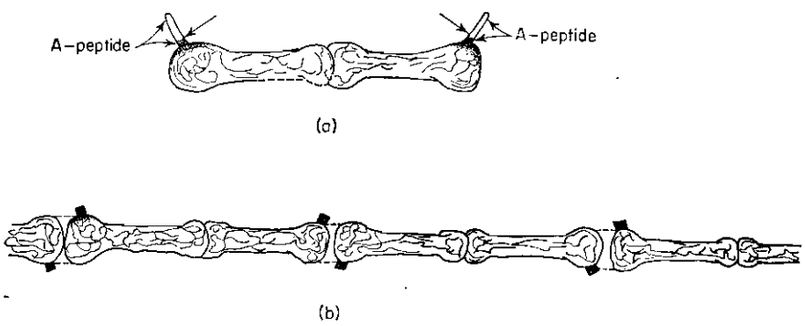


FIG. 1. Model of the fibrinogen molecule (a) and initial polymerization process (b).

has a narrow specificity of action on all mammalian fibrinogens so far investigated. (2) When cross-reacting thrombin and fibrinogen from several mammalian species, thrombin from different species shows remarkably small differences with regard to its clotting action on fibrinogen from one species (Fig. 3) (Blombäck and Teger-Nilsson, 1965). These facts indicate that at least the "active site" of thrombin has remained essentially unchanged in the different species. The fact that the amino acid sequence of fibrinopeptides in different species varies makes it likely that appreciable changes have occurred during evolution also in other parts of the fibrinogen molecule. However, the changes which have occurred in fibrinogen have apparently not been sufficient to induce co-adaptation of the "active site" of thrombin. It is therefore likely that the location of the cleavage site on the fibrinogen molecule has remained basically unchanged. If the cleavage site had

changed, one would expect on the basis of the narrow specificity of the enzyme that a change in the enzyme also had occurred.

This reasoning leads to the conclusion that the structures in the different species around the bond split are likely to represent genetically equivalent portions of the molecule. Comparative studies of the amino acid sequences of the fibrinopeptides from the cleavage site to the N-terminal ends of the peptides would thus be expected to give valuable taxonomic information not only with regard to the fibrinogen but perhaps also to the species. As the fibrinopeptides are easily isolated

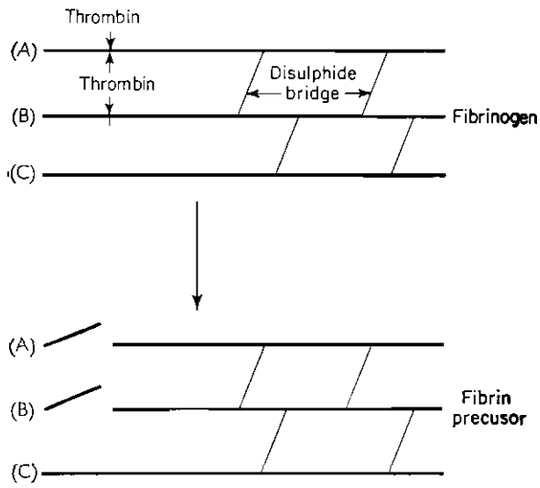


FIG. 2. Schematic representation of the chains in fibrinogen and fibrin.

and characterized, the fibrinogen is a particularly suitable protein for taxonomic studies.

In Fig. 4 are shown the amino acid sequences of fibrinopeptide A from different species. The amino acid sequences were mainly determined by Edman stepwise degradation. Usually only one main type of fibrinopeptide A or B can be demonstrated. In a hybrid, however, such as the mule, exactly equal amounts of horse and donkey chains are present. It is interesting to note that in the blood of one zebra roughly equal amounts of two A-peptides with different structures are present. Is this zebra a heterozygote of two fibrinogen alleles or is the zebra a fertile hybrid? It is evident that the residues close to the C-terminal end of the fibrinopeptide A have been rather stable during mammalian evolution. On the other hand, the N-terminal part of the chain has

changed much more. The C-terminal part of the chain, and especially positions 1 and 9, thus seems to have been subjected to a strong natural

FIBRINOGEN

Thrombin	Man	Monkey	Dog	Fox	Horse	Rat
Man	1·0	1·0	1·4	1·2	0·6	0·3
Monkey	1·0	0·9	1·3	1·1	0·5	0·3
Dog	0·9	0·9	1·3	1·1	0·5	0·3
Fox	1·0	0·9	1·4	1·2	0·6	0·6
Horse	0·9	0·9	1·1	0·9	0·6	0·2
Rat	1·0	0·9	1·3	1·1	0·5	0·3

Man–Monkey — Phe-Leu-Ala-Glu-Gly-Gly-Gly-Val-Arg
Dog–Fox — Phe-Ilu-Ala-Glu-Gly-Gly-Gly-Val-Arg
Horse — Phe-Leu-His-Glu-Gly-Gly-Gly-Val-Arg
Rat — Phe-Ilu (-Asp, Glu, Gly, Ala, Gly, Ilu, Arg)

FIG. 3. Clotting activity of thrombin from different species. The thrombins were assayed with bovine fibrinogen as reference. The activity of human thrombin on human fibrinogen is arbitrarily denoted as 1·0. Six other mammalian species showed essentially the same lack of species specificity with respect to thrombin. The amino acid sequences at the C-terminal ends of the fibrinopeptide A from the investigated species are included in the figure (Teger-Nilsson, unpublished).

selection and the structure in this part might, therefore, be of importance for a rapid association between enzyme and substrate or for the activation of the enzyme. We are now, on different grounds, of the opinion that a definite space relationship between position 1 arginine and position 9 phenylalanine is of particular importance for the action of the enzyme. A stable space relationship would be obtained if this part of the chain had a helical structure.

It is also evident that there is a striking resemblance in structure in the C-terminal portion not only between different eutherian mammals but also between marsupial and eutherian fibrinopeptide A. Even in a species of Saurian reptile the resemblance in structure is remarkable. It is to be noted that the fibrinogen from this lizard clotted fast and completely on addition of mammalian thrombin. This is not, however, a

19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1			
		H	-Ala	-Asp	-Ser	-Gly	-Glu	-Gly	-Asp	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Man	
		H	-Ala	-Asp	-Thr	-Gly	-Glu	-Gly	-Asp	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Green monkey (*C. aeth.*)	**Primates**
		H	-Ala	-Asp	-Thr	-Gly	-Glu	-Gly	-Asp	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Macaques (*Rh., Cyn.*); Baboon	
H	-Glu	-Asp	-Gly	-Ser	-Asp	-Pro	-Pro	-Ser	-Gly	-Asp	-Phe	-Leu	-Thr	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Ox	
H	-Glu	-Asp	-Gly	-Ser	-Asp	-Pro	-Ala	-Ser	-Gly	-Asp	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Bison (*B. bonasus*)	
H	-Ala	-Asp	-Gly	-Ser	-Asp	-Pro	-Ala	-Ser	-Ser	-Asp	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Red deer (*C. elaphus*)	
H	-Ala	-Asp	-Gly	(Ser,	Asp,	Pro,	Ala,	Ser,	Ser,	Glu,	Phe,	Leu,	Ala,	Glu,	Gly,	Gly,	Gly,	Val,	Arg)	-OH	Sika deer (*C. nippon*)	
H	-Ala	-Asp	-Gly	-Ser	-Asp	-Pro	-Ala	-Gly	-Gly	-Glu	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Reindeer, Elk (*A. alces*)	**Artiodactyls**
H	-Ala	-Asp	-Asp	-Ser	-Asp	-Pro	-Val	-Gly	-Gly	-Glu	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Sheep, Goat, *A. lervia, O. musimon*	
	H	-Ala	-Glu	-Val	-GlN	-Asp	-Lys	-Gly	-Glu	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Pig, Boar (*S. scrofa*)	
	H	-Thr	-Asp	-Pro	-Asp	-Ala	-Asp	-Lys	-Gly	-Glu	-Phe	-Leu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Llama	
	H	-Thr	-Asp	-Pro	-Asp	-Ala	-Asp	-Glu	-Gly	-Glu	-Phe	(Leu,	Ala,	Glu,	Gly,	Gly,	Gly,	Val,	Arg)	-OH	Camel (*C. drom.*	
					H	-Thr	-Glu	-Glu	-Gly	-Glu	-Phe	-Leu	-His	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Horse } Mule	
			H	-Thr	-Lys	-Thr	-Glu	-Glu	-Gly	-Glu	-Phe	-Ilu	-Ser	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Donkey }	**Perissodactyls**
			H	-Thr	-Lys	-Thr	-Glu	-Glu	-Gly	-Glu	-Phe	-Ilu	-Gly	-Glu	-Gly	-Gly	(Gly,	Val,	Arg)	-OH	Zebra	
			H	-Thr	-Lys	-Thr	-Glu	-Glu	-Gly	-Glu	-Phe	-Ilu	-Ser	-Glu	-Gly	(Ala,	Gly,	Val,	Arg)	-OH		
			H	-Thr	-AsN	-Ser	-Lys	-Glu	-Gly	-Glu	-Phe	-Ilu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Dog, Fox	
			H	-Thr	-Asp	-Gly	-Lys	-Glu	-Gly	-Glu	-Phe	-Ilu	(Ala,	Glu,	Gly,	Oly,	Gly,	Val,	Arg)	-OH	Bear (*U. arctos*)	
			H	-Gly	-Asp	-Val	-GlN	-Glu	-Gly	-Glu	-Phe	-Ilu	-Ala	-Glu	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Cat	
			H	-Thr	-Asp	-Thr	-Lys	-Glu	(Ser,	Asp,	Phe,	Leu,	Ala,	Glu,	Gly,	Gly,	Gly,	Val,	Arg)	-OH	Seal (*H. grypus*)	**Carnivores**
			H	-Thr	-Asp	-Val	-Lys	-Glu	-Ser	-Glu	-Phe	-Ilu	-Ala	-Glu	-Gly	-Ala	-Val	(-Gly | -Arg)	-OH	Badger (*M. meles*)	
			H	-Thr	-AsN	-Val	-Lys	-Glu	-Ser	-Glu	-Phe	-Ilu	-Ala	-Glu	-Gly	-Ala	(-Ala	-Gly	-Arg)	-OH	Mink (*M. vison*)	
		H	-Ala	-Asp	-Thr	-Gly	-Thr	-Thr	-Ser	-Glu	-Phe	-Ilu	(-Asp,	Glu,	Gly,	Ala,	Gly,	Ilu,	Arg)	-OH	Rat	Rodents
						H	-Thr	-Asp	-Thr	-Glu	-Phe	-Glu	-Ala	-Ala	-Gly	-Gly	-Gly	-Val	-Arg	-OH	Guinea pig	
		H	-Val	-Asp	-Pro	-Gly	-Glu	-Ser	-Thr	-Phe	-Ilu	-Asp	-Glu	-Gly	-Ala	-Thr	-Gly	-Arg	-OH		Rabbit	Lagomorphs
				H	-Thr	-Lys	-Asp	-Glu	-Gly	-Thr	-Phe	-Ilu	-Ala	-Glu	-Gly	-Gly	(Gly,	Val,	Arg)	-OH	Kangaroo	Marsupials
				H	-Thr	-Lys	-Thr	-Glu	-Gly	-Ser	-Phe	(Leu,	Ala,	Glu,	Gly,	Gly,	Gly,	Val,	Arg)	-OH	Wombat	
					H	-Glu	-Asp	-Thr	-Gly	-Thr	-Phe	-Glu	-Glu	-Gly	(His,	Gly,	Gly,	Val,	Arg)	-OH	Lizard (*T. rugosus*)	Sauria

FIG. 4. Fibrinopeptide A from different species. The sequences of baboon (determined by J. Hampton), elk, moufflon (*Ovis musimon*), maned sheep (*Ammotragus lervia*), boar, zebra, bear, seal, kangaroo, wombat and stumpy tail lizard (*Thrachydosaurus rugosus*) (determined by the authors together with C. Hann) have not been reported previously. The other sequences were reported by Blombäck *et al.* (1966). For further information see the latter paper.

general situation as several non-mammalian fibrinogens do not clot properly with mammalian thrombin, for instance lamprey fibrinogen (Doolittle, 1965).

Figure 5 shows the amino acid sequence of fibrinopeptide B from different species. Except for position 1, arginine, no area in the B-peptide is as invariable as the C-terminal part of the A-peptide. This might be due to a lesser force of natural selection acting on the B chain. It should be mentioned in this connection that the release of the B-peptide is slower than that of the A-peptide and its release is not of any importance in initiating the fibrinogen–fibrin transition. However, its role in the later process of polymerization has not yet been clarified.

The great variation in amino acid sequences in fibrinopeptides from different species shows clearly that they can be used as a tool for taxonomy and phylogeny. Seventeen of the 20 naturally occurring amino acids have been identified in different fibrinopeptides. The distribution of the 17 amino acids in the different positions in some of the fibrinopeptides is shown in Fig. 6.

Based on properties of the fibrinopeptides, sueh as (1) chain length, (2) point mutations and (3) sequence characteristics we can classify the different species (Blombäck *et al.*, 1966). The different Linnaean orders and suborders can evidently be well characterized by the properties of their fibrinopeptides. To give an example of the classification procedure we can take the subdivision of the artiodactyls as an example (Fig. 4). The ruminants are characterized by chain length of fibrinopeptide A of 19 residues, whereas Suiformes and Tylopoda are characterized by a chain length of 17 and 18, respectively. To give an example of a point mutation we can look at the llama and the camel. The two peptides are identical except for one residue. The llama has lysine in position A 12, whereas the camel has glutamic acid. As an example of a sequence characteristic we can take the Ser-Asp-Pro sequence which is typical for the ruminants.

As a further example of the great number of characteristics that can be found among these peptides we would like to draw your attention to fibrinopeptide B (Fig. 5). In the artiodactyls position 16 contains a tyrosine residue which is sulphated at the hydroxyl group. In the perissodactyls the tyrosine residue, also sulphated, is at position 17 (at least in the horse and donkey). In thc carnivores there is a duplication of the tyrosine. In the dog and fox these residues occupy positions 18 and 17, whereas in the cat they occupy positions 17 and 16. In the latter species both tyrosines are sulphated. In the dog group, however, two types of B-peptides are present, one where only one of the tyrosine residues is sulphated, and one where both are present as tyrosine-O-sulphate. In the kangaroo the corresponding tyrosine is at position 18 and is most

FIBRINOPEPTIDE B

21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1		
							[Glu	-Gly	-Val	-AsN	-Asp	-AsN	-Glu	-Glu	-Gly	-Phe	-Phe	-Ser	-Ala	-Arg	-OH	Man	
											H	-AsN	-Glu	-Glu	-Gly	-Leu	-Phe	-Gly	-Gly	-Arg	-OH	Green monkey (*C. aeth.*)	Primates
											H	-AsN	-Glu	-Glu	-Ser	-Pro	-Phe	-Ser	-Gly	-Arg	-OH	Macaques (*Rh., Cyn.*)	
[Glu	-Phe	-Pro	-Thr	-Asp	-Tyr	-Asp	-Glu	-Gly	-GlN	-Asp	-Asp	-Arg	-Pro	-Lys	-Val	-Gly	-Leu	-Gly	-Ala	-Arg	-OH	Ox	
(Glu,	Phe,	Pro,	Thr,	Asp,	Tyr,	Asp,	Glu,	Gly,	Glu,	-Asp,	Asp,	Arg,	Pro,	Lys)	-Val	-Gly	-Leu	-Gly	-Ala	-Arg	-OH	Bison	
(Glu,	Ser,	His,	Thr,	Asp,	Tyr,	Asp,	Glu,	Glu,	Glu,	Glu,	Asp,	Arg)	(Ala	-Lys)	(Leu,	His,	Leu,	Asp,	Ala,	Arg)	-OH	Red deer	
His	-Glu	-Leu	-Ala	-Asp	-Tyr	-Asp	-Glu	-Val	-Glu	-Asp	-Asp	-Arg	-Ala	-Lys	-Leu	-His	-Leu	-Asp	-Ala	-Arg	-OH	Reindeer	
H	-Gly	-Tyr	-Leu	-Asp	-Tyr	-Asp	-Glu	-Val	-Asp	-Asp	-AsN	-Arg	-Ala	-Lys	-Leu	-Pro	-Leu	-Asp	-Ala	-Arg	-OH	Sheep, Goat	Artiodactyls
	H	-Ala	-Ilu	-Asp	-Tyr	-Asp	-Glu	-Asp	-Glu	-Asp	-Gly	-Arg	-Pro	-Lys	-Val	-His	-Val	-Asp	-Ala	-Arg	-OH	Pig	
	H	-Ala	-Thr	-Asp	-Tyr	-Asp	-Glu	-Glu	-Glu	-Asp	-Asp	-Arg	-Val	(-Lys,	Val,	Arg,	Leu,	Asp,	Ala,	Arg)	-OH	Llama	
	H	-Ala	-Thr	-Asp	-Tyr	-Asp	-Glu	-Glu	-Glu	-Asp	-Asp	-Arg	-Val	-Lys	-Val	-Arg	-Leu	-Asp	-Ala	-Arg	-OH	Camel (*C. drom.*)	
	H	-Leu	-Asp	-Tyr	-Asp	-His	-Glu	-Glu	-Glu	-Asp	-Gly	-Arg	-Thr	-Lys	-Val	-Thr	-Phe	-Asp	-Ala	-Arg	-OH	Horse	
	H	-Leu	-Asp	-Tyr	-Asp	-His	-Glu	-Glu	-Glu	-Asp	-Gly	-Arg	(Thr	-Lys)	(Val,	Thr,	Phe	[Asp,	Ala,	Arg])	OH	Donkey	Perissodactyls
	H	-Leu	-Asp	-Tyr	-Asp	-His	-Glu	-Glu	-Glu	-Asp	-Gly	(Arg,	Ala,	Lys,	Val,	Thr,	Phe,	Asp,	Ala,	Arg)	-OH	Zebra	
	H	-His	-Tyr	-Tyr	-Asp	-Asp	-Thr	-Asp	-Glu	-Glu	-Glu	-Arg	-Ilu	-Val	-Ser	-Thr	-Val	-Asp	-Ala	-Arg	-OH	Dog	
		(Glu,	Tyr,	Tyr,	Asp,	Asp,	Thr,	Asp,	Glu,	Glu,	Glu,	Arg,	Ilu,	Val,	Ser,	Thr,	Val,	Asp,	Ala,	Arg)	-OH	Fox	Carnivores
H	-Ilu	-Ilu	-Asp	-Tyr	-Tyr	-Asp	-Glu	-Gly	-Glu	-Glu	-Asp	-Arg	-Asp	-Val	-Gly	-Val	-Val	-Asp	(-Ala, | Arg)	-OH	Cat	
							H	-Ala	-Thr	-Thr	-Asp	-Ser	-Asp	-Lys	-Val	-Asp	(Ilu,	*Ser,*	*Leu,*	*Ala,*	*Arg*)	-OH	Rat	*Rodents*
							H	-Ala	-Asp	-Asp	-Tyr	-Asp	-Asp	-Glu	-Val	-Leu	-Pro	-Asp	-Ala	-Arg	-OH	Rabbit	Lagomorphs
H	-Ser	-Phe	-Asp	-Tyr	-Asp	-Asp	-Asp	-Thr	-Gly	(Glu,	Gly,	Gly,	Gly,	Val,	Lys,	Ser,	Ser,	Val,	Asp,	Ala,	Arg)	-OH	Kangaroo	Marsupials

FIG. 5. Amino acid sequences of fibrinopeptide B. All sequences except those of zebra and kangaroo have been published previously (Blombäck *et al.*, 1966). The reader is referred to that publication for further information. The tyrosine residues at position 10 and 16–18 are generally sulphated (cf. Krajewski and Blombäck, 1968).

Position	A-peptide Amino acid residues	B-peptide Amino acid residues
1	Arg	Arg
2	Gly; Val	Ala; Gly
3	Ala; Gly; Thr; Val	Asp; Gly; Ser
4	Ala; Gly	Leu; Phe; Pro; Val
5	Gly	Arg; Gly; His; Leu; Phe; Pro; Thr; Val
6	Ala; Glu	Asp; Gly; Leu; Ser; Val
7	Ala; Asp; His; Ser; Thr	Glu; Lys; Val
8	Glu; Ilu; Leu	Ala; Asp; Glu; Ilu; Lys; Pro; Thr; Val
9	Phe	Arg; AsN; Asp
10	Asp; Glu; Thr	AsN; Asp; Glu; Gly; Ser; Tyr
11	Gly; Ser; Thr	AsN; Asp; Glu
12	Asp; Glu; Gly; Lys; Ser; Thr	Asp; GlN; Glu; Thr; Val
13	Ala; Asp; GlN; Glu; Gly; Lys; Pro; Thr; Val	Ala; Asp; Glu; Gly; Thr; Val
14	Ala; GlN; Gly; Pro; Ser; Thr; Val	Ala; GlN; Glu; Thr
15	AsN; Asp; Lys; Thr; Val	Asp; His
16	Ala; Asp; Glu; Gly; Pro; Ser; Thr; Val	Asp; Tyr
17	Ala; Asp; Gly	Asp; Tyr
18	Asp; Thr	Ala; Asp; Ilu; Leu; Thr; Tyr
19	Ala; Glu	Ala; GlN; His; Ilu; Leu; Pro; Tyr
20		Glu; Gly; Ilu; Phe
21		GlN; His

FIG. 6. Amino acid residues at different positions in fibrinopeptide A and B in mammals. (From Blombäck *et al.*, 1966.)

Pyroglutamic acid is counted as GlN. Positions 19 in fox and 21 in bison B-peptide are counted as GlN.

Red deer B-peptide sequence 10–21, rat A-peptide sequence 1–7 and rat B-peptide sequence 1–5 are not included in this table.

Comparison with sequence of	Man (Primates)			Ox (Artiodactyla)			Horse (Perissodactyla)			Dog (Carnivora)		
	A	B	A + B	A	B	A + B	A	B	A + B	A	B	A + B
Man	*100*	*100*	*100*	63	23	43	67	18	41	69	12	40
Green monkey	94	52	76	63	13	40	73	14	45	63	7	37
Macaques	94	52	76	63	7	37	73	14	45	63	14	40
Ox	63	23	43	*100*	*100*	*100*	55	35	44	46	20	32
Bison	69	23	46	90	100	95	55	35	44	51	20	34
Red deer	63	—	—	79	—	—	49	—	—	46	—	—
Sika deer	57	—	—	74	—	—	55	—	—	51	—	—
Reindeer	63	17	40	74	52	63	61	45	52	57	30	43
Sheep/goat	63	12	38	68	49	58	61	36	47	57	26	41
Pig	61	12	36	50	55	53	65	53	58	61	42	51
Llama	65	18	42	54	65	60	63	53	57	59	32	44
Camel	71	18	45	54	65	60	69	53	60	65	32	47
Horse	67	18	41	55	35	44	*100*	*100*	*100*	67	42	53
Donkey	56	18	37	46	35	40	80	100	90	75	42	57
Dog	69	12	40	46	20	32	67	42	53	*100*	*100*	*100*
Fox	69	12	40	46	20	32	67	42	53	100	95	97
Cat	69	29	48	51	39	45	67	41	52	75	51	62
Badger	44	—	—	29	—	—	40	—	—	63	—	—
Mink	38	—	—	23	—	—	40	—	—	69	—	—
Guinea-pig	48	—	—	38	—	—	52	—	—	55	—	—
Rabbit	44	22	34	34	24	29	33	31	32	38	19	28

FIG. 7. Degree of sequential correspondence between fibrinopeptides of various mammals. (From Blombäck *et al.*, 1966.) The figures denote in percentage the number of positions in which the compared sequences are identical. If the chain lengths are different, the percentage has been calculated on the average chain length.

likely also sulphated. In the rabbit the tyrosine sulphate has moved along the chain in direction towards the C-terminal part. The different location of the tyrosine residue in the different species is best explained by easy interconvertibility of tyrosine (tyrosine-O-sulphate) and aspartic acid in the fibrinopeptides. One can thus visualize how, by means of a series of reversible Asp to Tyr changes the tyrosine sulphate can move along the chain in either direction. Such a mechanism, rather than a deletion, explains the location of tyrosine sulphate at position 10 in the rabbit. In the primates and also rat the tyrosine sulphate is missing. It should be noted that in those species a considerable chain shortening has occurred.

With regard to taxonomy we can conclude that animals considered to be closely related within the orders show generally greater resemblances than those which according to the classical system are further apart. The correspondence in sequence, expressed as a percentage, is shown in Fig. 7. These figures do not, of course, give as much information as the weighting of the individual residues and sequences at specified positions in the chain. Some species which have been studied more recently are not included in the figure. Among the primates the baboon has the same sequence of the A-peptide as the other monkeys. The elk has an identical A-peptide to the reindeer, but its B-peptide differs at two positions from this animal. The overall agreement is thus 95% of reindeer. Of the two A-peptides found in the zebra both agree to 94% with the donkey, but only to 80 and 73%, respectively, with the horse. There is only one B-peptide in the zebra and this is, except for one residue, identical with both horse and donkey. The overall resemblance of the zebra peptides is definitely closer to the donkey than to the horse. Among the carnivores bear and seal show 88 and 69% correspondence, respectively, with the dog. Unfortunately, the B-peptides of these species have not yet been analyzed. The kangaroo A- and B-peptides show in comparison with the mammals high correspondences with the fibrinopeptides of the dog–cat group. There are also interesting homologies in the sequence between them. The stumpy tail lizard shows low correspondence with all compared species.

Although in general the homologies between the fibrinopeptides agree with generally accepted taxonomy, there are important exceptions from this rule observed, e.g. in the case of sheep and goat in comparison to the ox. According to the fibrinopeptide system sheep and goat are more closely related to the reindeer than to the ox. This is not in agreement with the findings of the morphological taxonomists.

Based on the sequence data phylogenetic trees for the different orders can be suggested. Figure 8 shows the phylogenetic relationship of the artiodactyls based on the structure of fibrinopeptides. Recent analysis

of the elk fibrinopeptides indicates that it should be on the reindeer line in this figure. This phylogenetic relationship is, except for the position of sheep and goat on the cervine line, in agreement with the view of organismal taxonomists. According to the classical concept, sheep and goat have emerged from the bovine line. Studies on the haemoglobin structure in the artiodactyls might explain this discrepancy. If the haemoglobin structure is in favour of the fibrinogen data, we believe classical morphologists have most probably been misled by anatomical convergence.

The phylogenetic 'tree' obtained for carnivores is shown in Fig. 9. From the limited data so far available one might suggest that the bear and the seal have also evolved from a dog-fox common line.

A suggestion of phylogenetic relationship between the primates is shown in Fig. 10. This relationship was deduced from the mutations, which have occurred at position 5 in fibrinopeptide B making use of the following code restriction Pro $\rightleftharpoons$ Leu $\rightleftharpoons$ Phe, which means that proline and phenylalanine cannot be directly converted into each other by one-base substitution in the corresponding RNA-code word. Of the two alternatives arrived at we have chosen that which agrees with morphological taxonomy (Alternative I).

Eck and Dayhoff (1966) have determined with the aid of computers the most likely phylogenetic relationship between the fibrinopeptides of a great number of the species and orders studied by us (Fig. 11). With regard to the groups we have already mentioned, their data agree well with our results.

With regard to the likely phylogenetic relationship between the perissodactyls, we would like to make a few comments. The two types of peptides in the zebra seem to be closely related to the donkey peptide A. All show striking differences from the horse peptide A. One may wonder if these species have undergone an evolution independent of the horse for a longer time period than is generally believed, or if domestication has accelerated mutational events in the horse. In this connection it is interesting to note that all changes between donkey and zebra can be explained by a one-base substitution in the corresponding RNA-code word. However, the horse peptide cannot be changed by one base substitution into either donkey or zebra peptides.

The basic event in an evolutionary process may be single amino acid changes by point mutations in a DNA codon. The agreement between the mutations which have occurred in the fibrinopeptides and a one-base substitution in the corresponding RNA-code words is shown in Fig. 12. As is seen here, the correspondence increases when closely related species are compared. Furthermore, when consideration is given to the most likely relationship of the species, as demonstrated in the

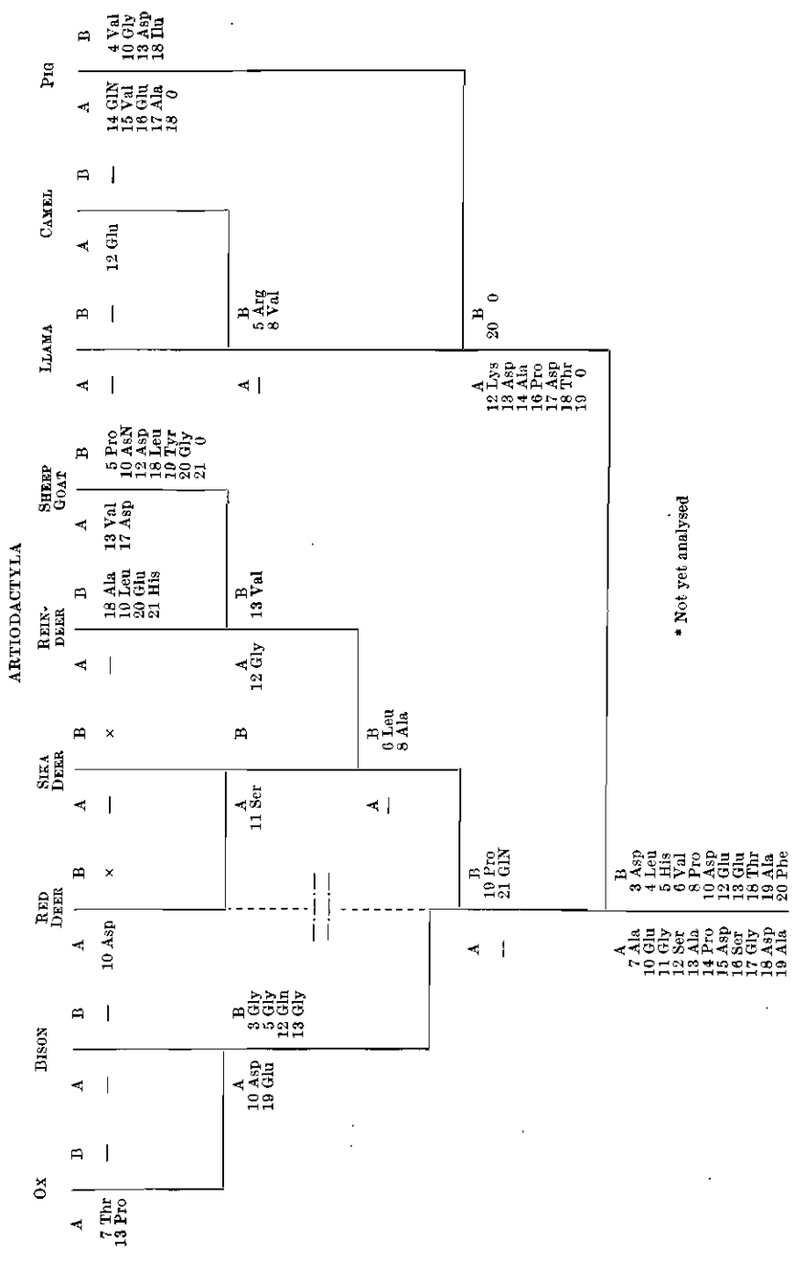


FIG. 8. Possible phylogenetic relationship of different artiodactyls. The relationship was obtained by comparison of residues at the same position in different species. (From Blombäck *et al.*, 1966.)

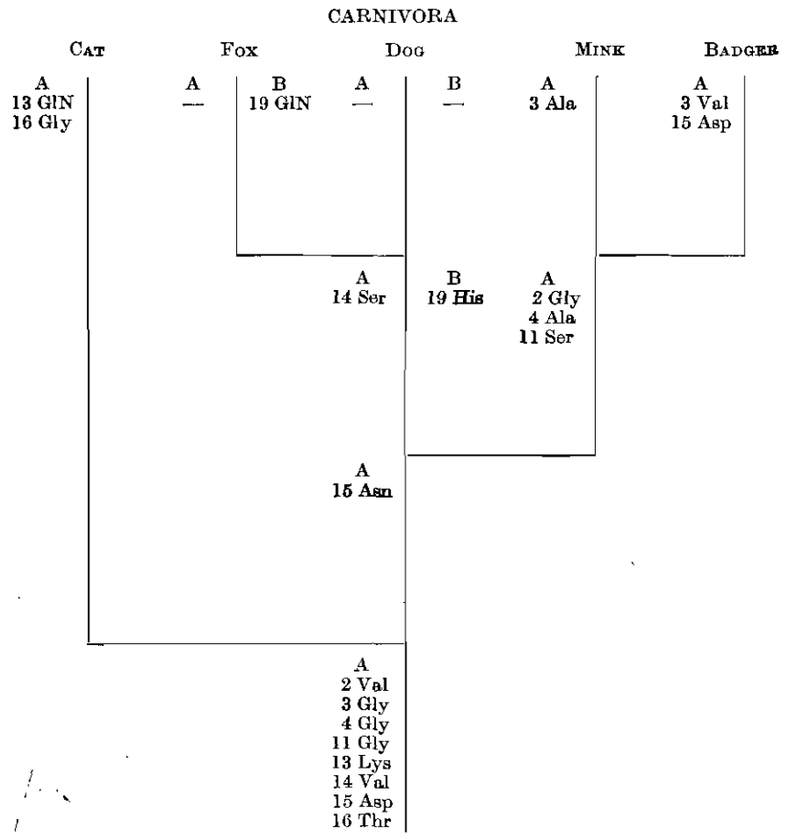


FIG. 9. Phylogenetic relationship of some carnivores. For construction of tree see Fig. 8. According to the sequential correspondence mink and badger could also have segregated as a separate line from the carnivore common stock. However, if consideration is given mainly to the residues A5–A16 the phylogenetic relationship shown in the figure is most likely. (*From Blombäck et al., 1966.*)

phylogenetic trees, a further restriction in possible mutational events is obtained, and the agreement with a one-base substitution in the RNA-code increases further to about 80%. These results indicate that the genetic RNA-code given by Nirenberg and associates may apply also to the mammalian protein system studied here.

We would like to finish our presentation with a discussion on the randomness of amino acid substitutions in the polypeptide chains of proteins. The fossil record, as we understand, indicates that the

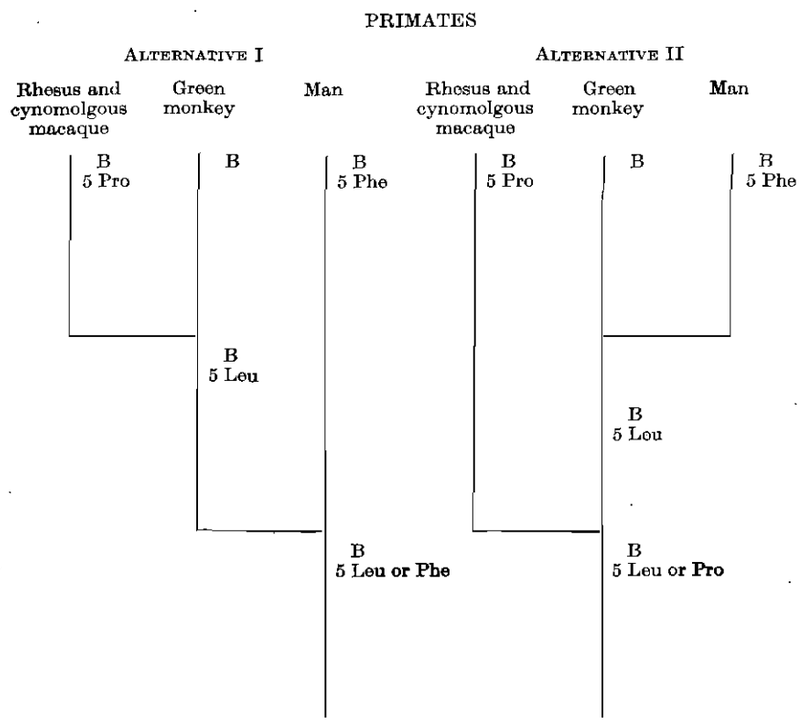


FIG. 10. Alternative phylogenetic relationship of some primates. (From Blombäck et al., 1966.)

emergence of species is confined to certain geological periods which indicates that the rate of change in the genetic material, and thus also in protein primary structure, has not been constant throughout all geological epochs. Therefore, on the basis of the surviving mutations observed in a protein from different species, we may at best suggest relative time scales for evolutionary events. The surviving amino acid substitutions we have described in the fibrinopeptides are apparently not random. We have suggested that any mutational event in the fibrinopeptide is primarily directed towards preserving definite tertiary structure in different parts of the chain. This means that a particular substitution may be deleted because it changes the conformation in a non-compatible way. However, it may survive if it is accompanied by compensatory secondary mutations in the vicinity of the particular primary mutation. This means that a single mutation may be followed by a burst of secondary mutations. If coupled mutations do occur,

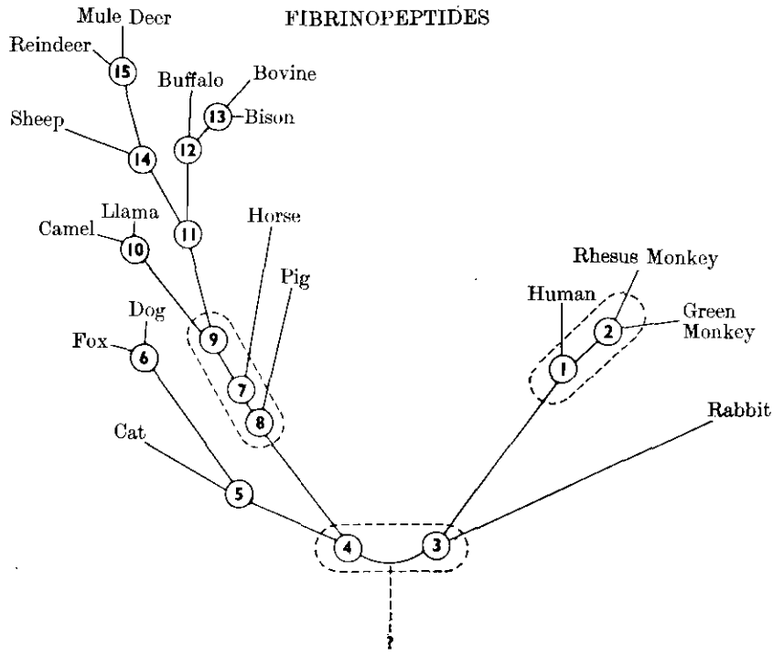


FIG. 11. Phylogenetic relationship of mammals. **(From Eck and Dayhoff, 1966.)**

	A-peptide		B-peptide	
	In all animals	In orders[a]	In all animals[b]	In orders[a]
Possible substitutions	286	76	354	114
Substitutions permissible with RNA-code	156 (55%)	54 (71%)	182 (51%)	74 (65%)

[a] In artiodactyls, carnivores and primates.
[b] Red deer B-peptide, rat A-peptide (step 1–7) and rat B-peptide (step 1–5) not included.

FIG. 12. Substitutions in fibrinopeptides and their agreement with one-base substitution in RNA-code. (From Blombäck *et al.*, 1966.)
Figures within brackets denote permissible substitutions in percentage of the possible substitutions

closely related species may show large differences in primary structure. Processes of this type could explain the discrepancy in the interpretation of fibrinopeptide and organismal data with regard to the classification of sheep and goat.

Evidence for protein conformation as the factor that eventually determines the selective advantage of mutational events has recently been put forward by Epstein (1967). Evidence also emerges from the comparative studies of rat and bovine ribonuclease (Beintema and Gruber, 1967). In these proteins about half of the amino acid residues are different. The peptide chain of rat ribonuclease can, even so, be folded into the same tertiary structure as that of the recently determined bovine ribonuclease.

REFERENCES

Beintema, J. J. and Gruber, M. (1967). Amino acid sequence of rat pancreatic ribonuclease. Abst. A-117, *Seventh Int. Congr. Biochem., Tokyo*, 1967.

Blombäck, B. (1967). Fibrinogen to fibrin transformation. *In* Seegers, W. H. (ed.) "Blood Clotting Enzymology", 143–215. Academic Press, New York.

Blombäck, B., Blombäck, M., Gröndahl, N. J. and Holmberg, E. (1966). Structure of fibrinopeptides—its relation to enzyme specificity and phylogeny and classification of species. *Arkiv Kemi*, **25**, 411–28.

Blombäck, B., Doolittle, R. F. and Blombäck, M. (1965). Fibrinogen: structure and evolution. *In* Peeters, H. (ed.) *XII. Coll. Protides Biol. Fluids, Bruges*, 1964, 87–94. Elsevier, Amsterdam.

Blombäck, B. and Teger-Nilsson, A.-C. (1965). On the thrombin–fibrinogen reaction in different species. *Acta chem. scand.* **19**, 751–3.

Doolittle, R. F. (1965). Characterization of lamprey fibrinopeptides. *Biochem. J.* **94**, 735–41.

Doolittle, R. F. and Blombäck, B. (1964). Amino acid sequence investigations of fibrinopeptides from various mammals: evolutionary implications. *Nature* **202**, 147–52.

Eck, R. V. and Dayhoff, M. O. (1966). "Atlas of Protein Sequence and Structure." National Biomedical Research Foundation, Maryland.

Epstein, C. J. (1967). Non-randomness of amino-acid changes in the evolution of homologous proteins. *Nature* **215**, 355–9.

Florkin, M. (1966). "A Molecular Approach to Phylogeny." Elsevier, Amsterdam.

Haschemeyer, A. E. V. (1963). A polar intermediate in the conversion of fibrinogen to fibrin monomer. *Biochemistry, N.Y.* **2**, 851–8.

Hill, R. L., Buettner-Janusch, J. and Buettner-Janusch, V. (1963). Evolution of hemoglobin in primates. *Proc. natn. Acad. Sci. U.S.A.*, **50**, 885–993.

Hill, R. L. and Buettner-Janusch, J. (1965). Evolution of hemoglobin in primates. *In* Peeters, H. (ed.) *XII. Coll. Protides Biol. Fluids, Bruges*, 1964, 95–101. Elsevier, Amsterdam.

Krajewski T. and Blombäck, B. (1968). The location of tyrosine-O-sulphate in fibrinopeptides, *Acta Chem. Scand.* (in press).

Margoliash, E. (1963). Primary structure and evolution of cytochrome *c*. *Proc. natn. Acad. Sci. U.S.A.* **50**, 672–9.

Margoliash, E., Needleman, S. B. and Stewart, J. W. (1963). A comparison of the amino acid sequences of the cytochromes c of the several vertebrates. *Acta chem. scand.* **17**, suppl. 1, 250–56.

Pauling, L. and Zuckerkandl, E. (1963). Chemical paleogenetics molecular "restoration studies" of extinct forms of life. *Acta chem. scand.* **17**, suppl. 1, 9–16.

2. Haemoglobin in Man and Animals

R. G. HUNTSMAN

Department of Pathology, Lambeth Hospital, London, England

ABSTRACT

Man's haemoglobin has been intensively studied and its intricate chemical structure is now known. As well as the normal haemoglobin, about 100 variant haemoglobins have been described, the great majority differing from the normal by the substitution of only a single abnormal amino acid.

The total analysis of a haemoglobin molecule is time consuming in the extreme and accounts in great part for the present ignorance of the chemical structure of animal haemoglobins. Preliminary work on a few species suggests possible correlation between the time that two species separated and the number of amino acid differences in their haemoglobins.

Human haemoglobin variants have proved of value as "racial markers" and explanations are offered to explain the unusual findings of a recent haemoglobin survey in this country.

It is suggested that a discussion on haemoglobin classification will be possible only when new information concerning the structure of animal haemoglobins becomes available.

DISTRIBUTION OF HAEMOGLOBIN

Haemoglobin occurs widely in nature, being, for instance, found in the root nodules of plants and the nerve cord of a marine invertebrate (*Aphrodite*). In the latter case Schafer (1898) described the colour of the nerve ganglia as "intense as human blood".

In vertebrates, haemoglobin-containing red cells are almost universally found in every species. Exceptions are the "ice fish" (Chaenichthyidae) which are confined to the Antarctic and can live without haemoglobin or red cells in their peripheral blood. They appear to owe their survival to their sluggish existence, within the cold well-aerated water of the Antarctic in which they swim (Ruud, 1954).

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 21–28.

For about 100 years it had been realized that haemoglobins of different species were different. Körber (1866) demonstrated that the rate of alkali denaturation of haemoglobin was variable in selected animals and showed also that the human foetus possessed different haemoglobin from the human adult. Funke (1851) showed that different haemoglobin from different species formed different crystals, and following this approach Reichert and Brown (1907) were able to appreciate that a blood sample of baboon must have been wrongly labelled and correctly identified it as cat! They appreciated that the differences in crystalline form represented chemical differences but one has had to wait 50 years for the analytical chemist to perfect his techniques to confirm this.

THE STRUCTURE OF THE MAMMALIAN HAEMOGLOBIN MOLECULE

Little is known about the chemical structure of invertebrate haemoglobins. The molecular weight is known to be highly variable and in some species may be as high as 3 000 000 (Keilin and Hartree, 1951). This section is concerned with the structure of mammalian haemoglobin only because it has been most intensely studied.

The haemoglobins are a group of related proteins (globins), to each of which the same prosthetic group, haem, is attached. The haem remains constant in all haemoglobins. In order for it to function effectively as a respiratory pigment, the iron atom must remain in the ferrous (Fe^{++}) state. Permanent oxidation of the ferrous atom to the ferric (Fe^{+++}) form is remarkably retarded by the deep insertion of the haem group into a pocket within the globin molecule. Within the pocket the haem is suspended between two histidine amino acid residues and the pocket itself is lined by amino acids with uncharged hydrophobic side chains, as if to prevent the iron from "rusting". If an amino acid residue with a charged side chain is, as a result of a mutation, inserted into the lining of the haem pocket, the iron atom may be permanently converted to the ferric form. This converts haemoglobin to methaemoglobin which is unable to carry molecular oxygen.

Myoglobin exists as a single haem/globin unit and this has a hyperbolic oxygen dissociation curve. The haemoglobin molecules of vertebrates, with the exception of the lamprey (Rumen, 1966), are predominantly in the form of a tetramer which gives the molecule a more advantageous sigmoid-shaped oxygen dissociation curve. The four sub-units of the haemoglobin molecule move closer together when the molecule is in the oxygenated state, and when the oxygen molecule leaves the haemoglobin the haemoglobin molecule expands. This expansion of the molecule permits histidine amino acid residues, which

were previously hidden, to become exposed and give reduced haemoglobin a greater buffering potential.

The globin is composed of amino acid chains, about 80% of which are in a helical form. These globin units, whether single as in the myoglobin of the sperm whale or part of a tetramer as in the α chain of human haemoglobin, have a certain essential fixed form which appears necessary for the efficient functioning of the molecule. This is achieved, although there are 107 amino acid differences in the 141 molecular sites common to sperm whale myoglobin and human α chains (Zuckerkandl and Pauling, 1965). Provided this fixed form is maintained, it is obvious that many amino acid substitutions are permissible in the polypeptide chain. For instance, of the approximately 150 amino acids making up a globin unit of the haemoglobin molecule there are only 9 sites (and this number may yet be reduced further) where the same amino acid occurs in the haemoglobin of all the species so far examined (Jukes, 1966).

In the haemoglobin molecule, the amino acids with charged side chains point towards the exterior (Perutz *et al.*, 1965), and this may help to make the molecule soluble. The internal bulk of the molecule appears to be composed of uncharged amino acids and the haem pocket is a particularly hydrophobic part. It is probable that many amino acid mutations, for example, the substitution of an uncharged for a highly charged amino acid in the interior of the molecule, may result in total disruption of the haemoglobin molecule, which would be incompatible with life. Certain amino acid substitutions in man, whilst not lethal, result in a haemoglobin molecule which is peculiarly unstable, the patient presenting with a persistent anaemia (Carrell *et al.*, 1967). Other amino acid substitutions, particularly involving the histidyl amino acid residues between which the iron is slung, result in haemoglobin which can only exist in the methaemoglobin form (the M haemoglobins) and which cannot function as an oxygen carrier.

It is only by the most careful and time-consuming chemical examination of haemoglobins found not only as haemoglobin variants in man but also as normal haemoglobins in animals, that the "rules" governing the positioning of amino acids and the function of the haemoglobin molecule will be thoroughly understood. It is, for instance, interesting that one of the few amino acid differences that distinguishes the haemoglobin of an orang utan from man also occurs in a rare human haemoglobin variant, Haemoglobin Norfolk (Buettner-Janusch and Hill, 1965).

CHEMICAL STRUCTURE OF ANIMAL HAEMOGLOBINS

The complete analysis of a haemoglobin molecule is immensely time-consuming and because of this the haemoglobin from comparatively

few species has been completely examined (Schroeder and Jones, 1965). It is therefore too early to form far-reaching conclusions as to the potential of the haemoglobin molecule as a species marker. It does, however, appear that the further apart species are the more dissimilar one finds their haemoglobin molecules. There are multiple amino acid differences between the polypeptide chains of carp and man (Rudloff *et al.*, 1966), about 20 between horse and man (Braunitzer *et al.*, 1964), and only 3 between gorilla and man (Zuckerkandl and Pauling, 1962). Taken in conjunction with other proteins and given enough chemical information, a chemical classification will have future value. It is because the main emphasis has been on the chemical analysis of the human haemoglobin variants that it is necessary to consider these rather than animal haemoglobins in some detail. The human haemoglobin variants have certainly proved of value in anthropology as racial markers.

THE STRUCTURE OF THE PHYSIOLOGICAL HUMAN HAEMOGLOBINS

Adult haemoglobin (Haemoglobin A) contains 4 polypeptide chains. Each molecule has two chains called α and two chains called β. Foetal haemoglobin (Haemoglobin F) which comprises about 70% of the haemoglobin of the new-born and which virtually disappears by two years, has γ chains in place of the β chains found in Haemoglobin A. Haemoglobin A_2, found at a level of only 1–2% in normal adults, appears to be present only in primates (Kunkel *et al.*, 1957), and has in place of the β and γ chains different chains called δ chains. The structure of the three physiological human haemoglobins may thus be summarized as follows:

$$\text{Haemoglobin A} = \alpha_2\beta_2$$
$$\text{Haemoglobin F} = \alpha_2\gamma_2$$
$$\text{Haemoglobin A}_2 = \alpha_2\delta_2$$

When the haemoglobins of various primates are examined, the different α chains appear to be more chemically alike than the different β chains. This is possibly explained by the fact that the α chain mutation would have to be able to pair off with both the β and γ chains to form adult and foetal haemoglobin. Amino acid substitutions in the β chain would only need to fit one partner structurally, the α chain (Hill *et al.*, 1963).

By the introduction of gaps in the polypeptide chains (probably representing previous deletions) it becomes possible to recognize that these polypeptide chains have certain similarities; indeed it has been suggested that they and myoglobin may well have originated from a common gene at some distant date in the past. Zuckerkandl and

Pauling (1962) have attempted to produce a time scale to show when the now separate genes for these various polypeptide chains originally split away from each other. They based their time scale on the fact that the horse and man probably had a common stock about 100–160 million years ago. There are about 18 amino acid differences between the α polypeptide chains of the haemoglobin of man and horse. In other words, since they parted from a common origin, these two genes have each undergone an average of 9 mutations, i.e. 1 mutation every 15 million years. This interesting speculation would place the parting of myoglobin and haemoglobin at about 600 million years ago, before the origin of the vertebrates!

HUMAN HAEMOGLOBIN VARIANTS

Apart from the M haemoglobins, which present with cyanosis because of methaemoglobinaemia, and the unstable haemoglobins which present as congenital anaemias, the heterozygote state for an abnormal haemoglobin is symptomless. These people are recognized as a result of surveys using electrophoresis, when the protein molecules are allowed to travel in a buffer in an electrical field. This technique unfortunately limits the identification of variants to those where a charge change has resulted from the amino acid substitution that has occurred. It must be recognized that any survey relying only on electrophoresis will in fact be missing the majority of mutations (for example a neutral amino acid to another neutral amino acid) that have occurred.

In England there has recently been a survey of 15 000 blood samples obtained from the umbilical cord of new-born babies. The main purpose of the survey was to assess the frequency of γ chain variants of foetal haemoglobin and four such variants have been found. One of these is of particular interest because it is a glutamic acid $\rightarrow$ lysine mutation in the 121st position of the γ chain, the polypeptide chain found only in foetal haemoglobin (Sacker *et al.*, 1967). A glutamic acid $\rightarrow$ lysine mutation has previously been described in the α chain in the adult variant O Indonesia and in the β chain in the adult variant O Arab—both at similar positions to that in the recently found γ chain haemoglobin variant. It is also of interest that a glutamic acid $\rightarrow$ glutamine mutation found in an adult variant Haemoglobin D Punjab occurs in the β chain at this same site. There is at the moment no satisfactory explanation to explain these numerous amino acid substitutions occurring at the same position in different polypeptide chains.

During this cord blood survey it was perhaps surprising that Haemoglobin S (commonly occurring in the Negroes) was found in a native British family. Also a number of examples of Haemoglobin D Punjab

(found mainly in Northern India) were also found in English and Irish families.

HAEMOGLOBIN S

It is possible that Haemoglobin S (sickle haemoglobin) originated in the Middle East and spread from there to the "sickle belt" of Africa. The presence of Haemoglobin S in the New World is due to the earlier transport of Negroes across the Atlantic. It is perhaps at first strange that Haemoglobin S is found at all in so-called native English, but the occasional findings of this pigment might well be explained by for instance the large number of Negro servants in this country in the 18th century (Hecht, 1954). Dr. Johnson had a Negro servant who carried the "Empire of Cupid further than most men". If the sexual activity of Dr. Johnson's servant was in any way average there is no doubt that their genes would have been fairly liberally sprinkled throughout the English community of that time.

HAEMOGLOBIN D PUNJAB

The geographical distribution of Haemoglobin D is perhaps the most interesting of all the common haemoglobin variants. It is centred in the Punjab where it occurs at a level of about 3%. It is also commonly found (about 1 in 2 000) amongst native English and Irish. For instance, it was picked up in an English family in a survey recently carried out in Norfolk. It appears probable that the presence of this pigment in the English merely confirms the bonds that existed between England and India during the last 300 years. Evidence based on one regiment (105th Madras Second European Regiment) suggests that there was a total of 400 marriages per annum between British troops and Indo-British women (Konigsberg *et al.*, 1965). It is of interest that the Norfolk regiment was stationed in India on a number of occasions and campaigned valiantly in the Punjab during the first Sikh war.

HAEMOGLOBIN BART'S

During the cord blood survey that was previously mentioned, a large number of examples of Haemoglobin Bart's (composed of four γ chains) was found in babies of recent immigrants from Asia and Africa. Only one example was found in a baby of a native English mother. On further investigation it was found that this mother was an unmarried mother and the baby had been conceived during the time that she was working in a Chinese restaurant! (Barakat *et al.*, 1967).

The above examples are given to illustrate some of the problems that arose following the finding of haemoglobin variants during a recent survey in England. I hope they have served to give the reader some idea of the type of discussions that will result when, after further chemical analysis has been done on the blood of various species of animals, an attempt at a "*haemoglobin classification*" commences.

REFERENCES

Barakat, A., Marengo-Rowe, A. J., Gaffney, P. J., Huntsman, R. G. and Lehmann, H. (1967). *Haemoglobin O Arab in Egypt and Aden*: Possible errors resulting from the use of haemoglobin variants as genetic markers in population surveys. *Z. Morph. Anthrop.* **59**, 100–103.

Braunitzer, G., Hilse, K., Rudloff, V. and Hilschmann, N. (1964). The Hemoglobins. *Adv. Protein Chem.* **19**, 1–71.

Buettner-Janusch, J. and Hill, R. L. (1965). Molecules and monkeys. *Science, N.Y.* **147**, 837–42.

Carrell, R. W., Lehmann, H., Lorkin, P. A., Raik, Eva and Hunter, Elizabeth (1967). Haemoglobin Sydney. (β67 (E11) Valine → Alanine): an emerging pattern of unstable haemoglobins. *Nature, Lond.* **215**, 626–8.

Funke, O. (1851). Ueber das Milzvenblut. *Z. Rat. Med.* **1**, 172–218.

Hecht, J. (1954). Continental and colonial servants in eighteenth century England. **40**, Smith College studies in history, Northampton, Massachusetts.

Hill, R. L., Buettner-Janusch, J. and Buettner-Janusch, V. (1963). Evolution of hemoglobin in primates. *Proc. natn. Acad. Sci.* **50**, 885–93.

Jukes, T. H. (1966). "*Molecules and Evolution.*" Columbia University Press, New York and London.

Keilin, D. and Hartree, E. F. (1951). Relationship between haemoglobin and erythrocruorin. *Nature, Lond.* **168**, 266–9.

Konigsberg, W., Huntsman, R. G., Wadia, F. and Lehmann, H. (1965). Haemoglobin D Punjab in an East Anglian family. *Jl. R. anthrop. Inst.* **95** (2), 295–306.

Körber, E. (1866). Ueber Differenzen des Blutfarbstoffes. Thesis for Medical Doctorate, Dorpat.

Kunkel, H. G., Ceppellini, R., Muller-Eberhard, U. and Wolf, J. (1957). Observations on the minor basic haemoglobin components in the blood of normal individuals and patients with thalassaemia. *J. clin. Invest.* **36**, 1615–25.

Perutz, M. F., Kendrew, J. C. and Watson, H. C. (1965). Structure and function of haemoglobin. *J. molec. Biol.* **13**, 669–78.

Reichert, E. T. and Brown, A. P. (1907). The crystallography of hemoglobins. *Proc. Soc. exp. Biol. Med.* **5**, 66–68.

Rudloff, V., Hilse, K., Zelenik, M. and Braunitzer, G. (1966). Vergleichende Untersuchungen der Primärstruktur verschiedener Hämoglobine. *Int. Symp. Comp. Haemoglobin Structure*, 149, Thessaloniki.

Rumen, N. M. (1966). A comparison of sea lamprey (*Petromyzon marinus*) and mammalian haemoglobins. *Int. Symp. Comp. Haemoglobin Structure*, 134, Thessaloniki.

Ruud, J. T. (1954). Vertebrates without erythrocytes and blood pigment. *Nature, Lond.* **173**, 848–50.

Sacker, L. S., Beale, D., Black, A. J., Huntsman, R. G., Lehmann, H. and Lorkin, P. A. (1967). Haemoglobin F. Hull (γ121 Glutamic acid → Lysine) homologous with haemoglobins O Arab and O Indonesia. *Br. med. J.* **3**, 531–33.

Schafer, E. A. (1898). "Textbook of Physiology." Young J. Pentland, Edinburgh and London.

Schroeder, W. A. and Jones, R. T. (1965). Some aspects of the chemistry and function of human and animal hemoglobins. *Fortschr. Chem. org. NatStoffe* **23**, 113–94.

Zuckerkandl, E. and Pauling, L. (1962). Molecular disease, evolution and genic heterogeneity. "Horizons in Biochemistry", 189. Academic Press, New York and London.

Zuckerkandl, E. and Pauling, L. (1965). Divergence and convergence in proteins. In Bryson, V., and Vogel, H. J. (eds.) "Evolving Genes and Proteins". Academic Press, New York and London.

3. Indoles, Amino Acids and Imidazoles as an Aid to Primate Taxonomy

IVOR SMITH

Courtauld Institute of Biochemistry,
The Middlesex Hospital Medical School,
London, England

ABSTRACT

Analysis by paper chromatography-electrophoresis of urine and blood specimens of primates shows that distinctive species-specific patterns exist. In blood these appear as quantitative changes whereas in urine major qualitative differences appear, particularly in the indoles but also in the amino acids and imidazoles.

Amino acids are the basic structural units of proteins and enzymes but, whereas these latter proteins have been investigated successfully with the aim of constructing a primate biochemical taxonomy, the amino acids themselves have been, with two exceptions (Gartler *et al.*, 1956; Fooden, 1961), ignored. Nevertheless, it would seem that the circulating levels of these amino acids together with their excretory metabolites should be equally important from the taxonomic viewpoint. I report here a summary of the results of a survey of these compounds in the blood and urine of a range of primates.

COLLECTION OF SPECIMENS

All specimens were collected from colonies within the United States. Specimens were obtained from the various regional and other primate centres and were of three types, namely: (1) dried blood spots—sent by airmail; (2) whole serum—sent by air in dry-ice containers; (3) urine—random specimens—also sent frozen.

Altogether blood spots and/or serum were obtained from 14 species

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 29–37.

and urine from 17 species. Names quoted in the text are those recorded by the donor laboratory: *Homo sapiens* (man), *Gorilla* sp. (gorilla), *Papio* sp. (baboon), *Pan* sp. (chimpanzee), *Macaca mulatta* (macaque or rhesus), *Hapale jacchus* = *Callithrix* (cotton-ear marmoset), *Galago* sp. (bush baby), *Tupaia glis* (tree shrew).

BLOOD AND SERUM

Blood spots, or serum spots similarly deposited on paper, were pretreated as described by Efron *et al.* (1964). Each spot, equivalent to 8–10 μl blood, was then inserted into a sheet of paper and run by two-way electrophoresis followed by chromatography in the Unikit apparatus (Smith and Feinberg, 1965) as described by Smith *et al.* (1967). In some cases, one-way chromatographic runs were also compared. Location reagents used were ninhydrin, Ehrlich's reagent and diazotized sulphanilic acid (Pauly) as previously described (Smith, 1968).

The blood spot amino acid pattern of the normal human neonate (Efron *et al.*, 1964) is remarkably constant following one-way paper chromatography. I have found that the pattern of the adult human and of each species examined is, likewise, constant on both one- and two-way chromatograms. As far as can be observed, the same qualitative blood amino acid pattern is obtained but relative quantitative levels of these amino acids are different.

Thus each species produces a characteristic quantitative pattern which is recognizably different from that of another species. Wherever possible, both blood and serum of the same species were compared and these yielded identical patterns but for the appearance of a single additional spot from blood in every case. The position of this spot leads to a tentative identification as red-cell glutathione. A representative selection of patterns is shown in Fig. 1. These chromatograms showed no indoles and no imidazoles except for varying amounts of histidine.

URINE

It is possible to collect 24-h specimens from most primates given sufficient time and freedom from other duties. However, a wealth of experience with human urine (Smith, 1968) shows that whereas the relative amounts of any given component may vary between random specimens the general qualitative pattern remains unchanged; thus, a random specimen never gives rise to an abnormal pattern. This was confirmed by comparison of 24-h and random specimens obtained from a number of different marmoset and other species in our own colony. Further, as the normal *Hapale jacchus* (marmoset) pattern was

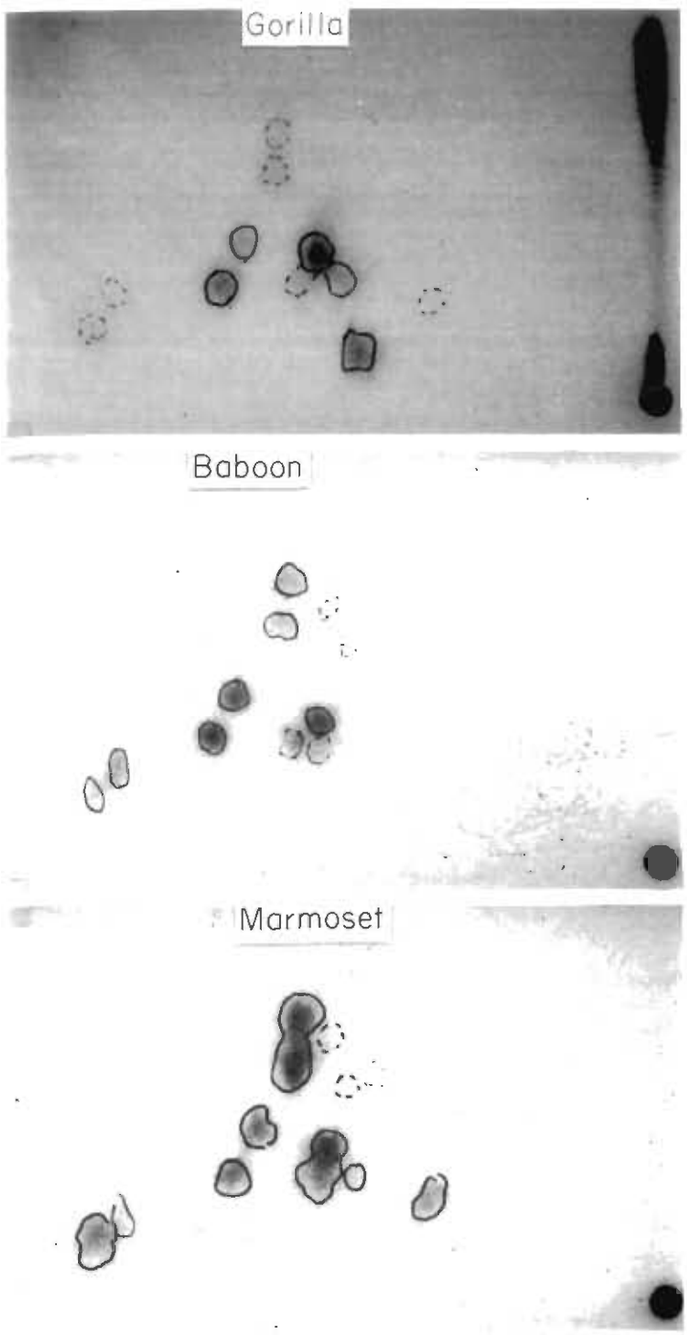


FIG. 1. Amino acids present in gorilla (*Gorilla* sp.) blood, and baboon (*Papio* sp.) and marmoset (*Hapale jacchus*) serum. Note the qualitative similarity and relative quantitative differences of the spots; also glutathione spot, shown square, present in blood but absent from serum.

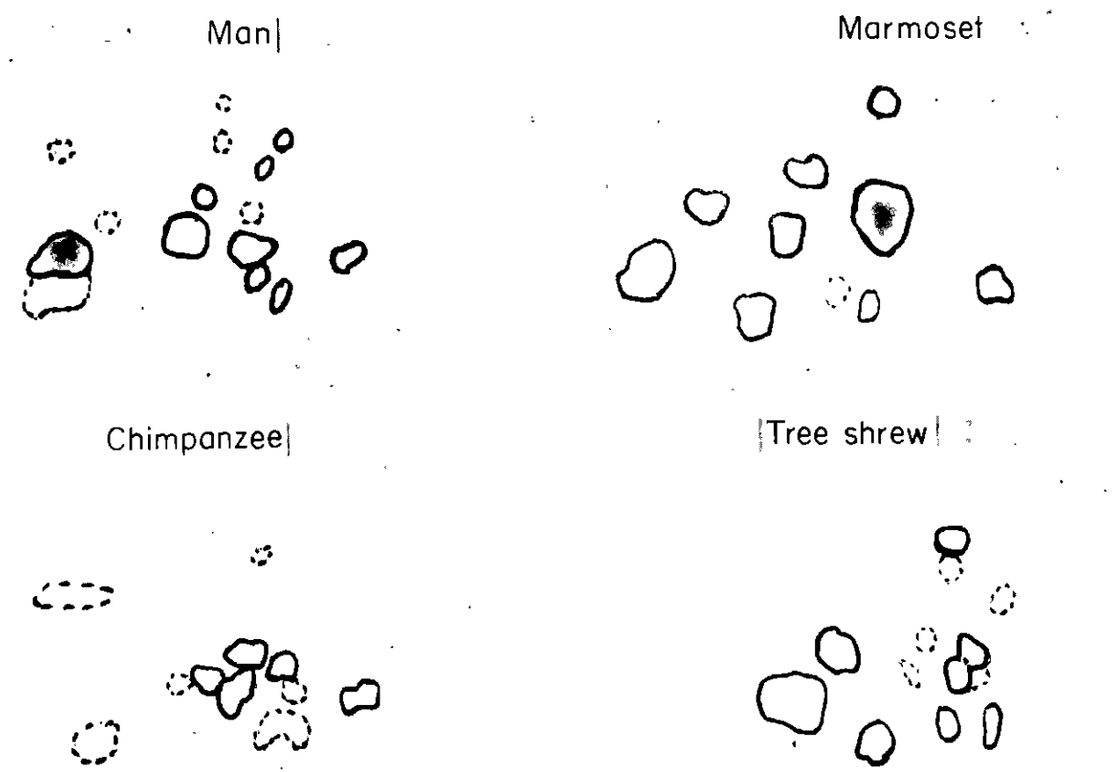


FIG. 2. Urinary amino acid patterns of man (*Homo sapiens*), chimpanzee (*Pan* sp.), marmoset (*Hapale jacchus*) and tree shrew (*Tupaia glis*) showing appreciable qualitative differences. Electrophoresis, pH 2, first in horizontal direction followed by chromatography.

qualitatively radically different in a number of respects from that of the human, it was confidently assumed that random specimens obtained from other species would show qualitative similarity within the species whilst retaining qualitative differences between the species.

Amino acids and imidazoles were located on the same sheet following the two-way electrophoresis–chromatography procedure discussed above. Indoles were located following two-way paper chromatography as described by Jepson (1962). The electrophoresis–chromatography procedure is equally valuable for the indoles but was not used here because all suitable apparatus was fully employed with the other compounds.

Qualitative differences between the amino acid patterns of many of the species is readily apparent. Further work with larger numbers of specimens from each species will, no doubt, confirm that they show different and characteristic "normal" patterns; a few examples of such patterns are shown in Fig. 2. The amino acids present in the urine represent the end products of metabolism plus the overflow of amino acids derived from the total ingested protein. An interpretation of the significance of these differences is therefore extremely difficult. Hence it was thought worth while to endeavour to follow the metabolic products derived from two particular amino acids, namely histidine and tryptophan, whose characteristic chemical reactions allow them, and many of their metabolites, to be readily differentiated in the presence of other amino acids.

Imidazoles are, presumably, all derived from histidine. After chromatographic separation, the only positive diazoreactor in man is *histidine*, although *in tropical climates imidazole lactic acid is* also frequently observed (Smith and Rider, 1967). Following electrophoresis, however, traces of imidazole lactic, propionic and acetic acids are sometimes observed. Other primates show different patterns which may, or may not, include a histidine spot. The interpretation is rendered difficult due to the large number of phenolic spots which show up with the diazo reagent but the imidazole nature of many of these may be confirmed by re-running after ethyl acetate extraction of the acidified specimen. Figure 3 shows a number of patterns obtained; the ringed spots represent true imidazoles whereas the others may be taken to be phenols, including hydroxy-indoles and phenolic amines.

Indoles are all derived from dietary tryptophan. Differences in indole patterns (i.e. Ehrlich-positive reactors), were by far the most striking. Every one of the 17 different species examined gave a characteristic pattern which enabled it to be distinguished from the other species; for examples see *Fig. 4*. Patterns ranged from no indoles in man (urea and indoxyl sulphate serve as valuable markers) up to as many as ten in one

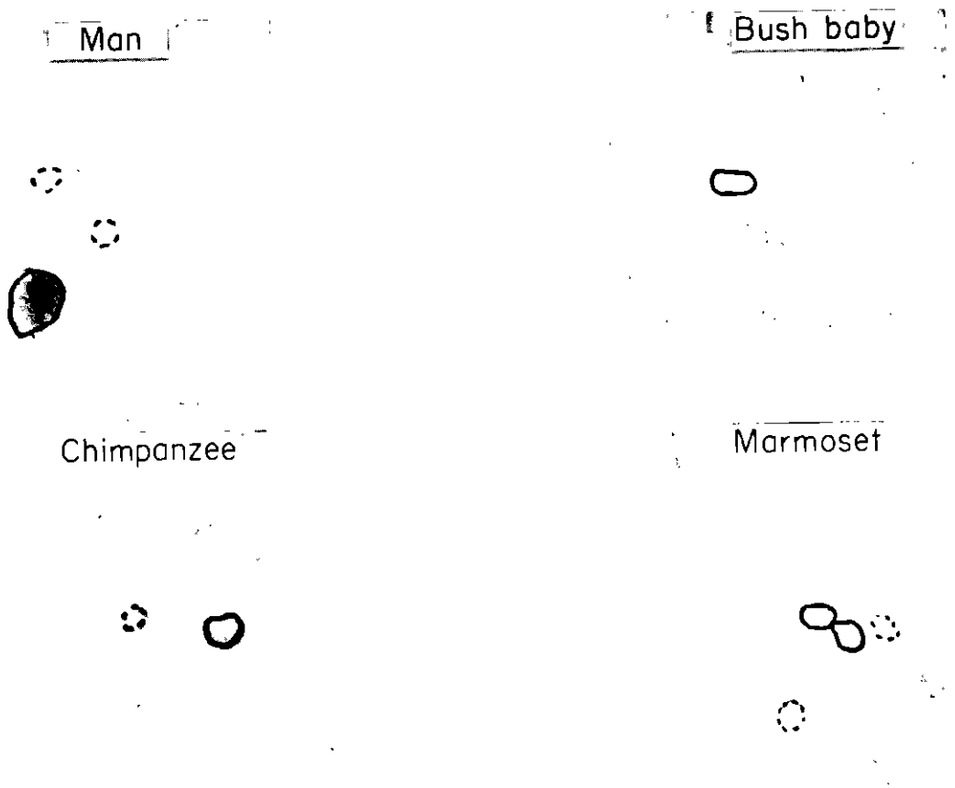


FIG. 3. Urinary imidazole patterns of man (*Homo sapiens*), chimpanzee (*Pan* sp.), bush baby (*Galago* sp.) and marmoset (*Hapale jacchus*). Ringed spots represent imidazoles, others are phenols. Separation conditions as for Fig. 2.

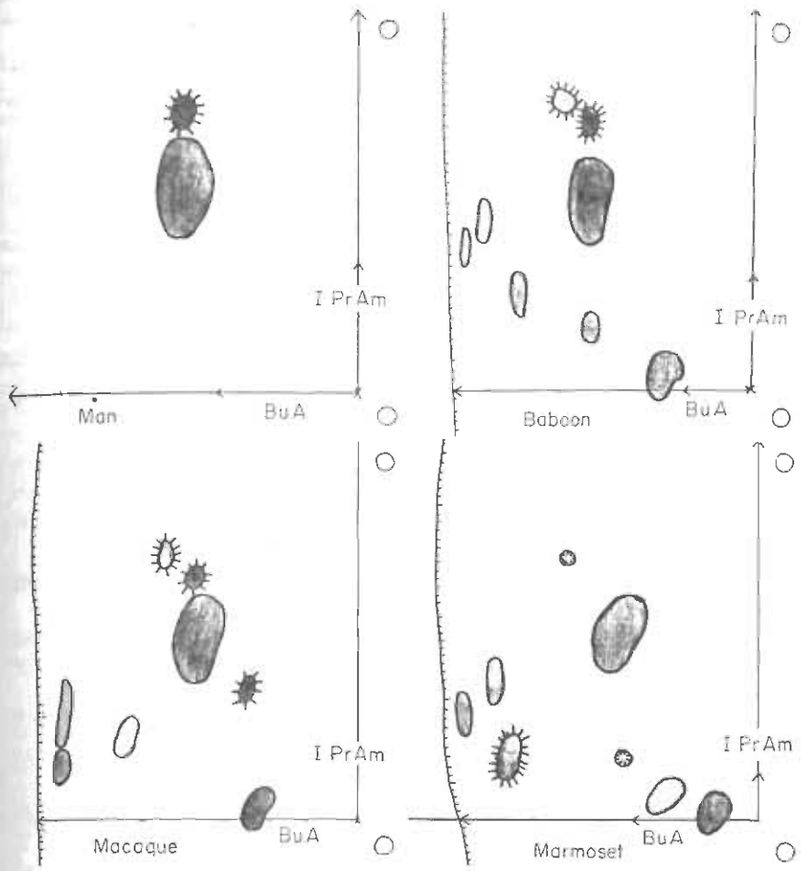


FIG. 4. Urinary indole patterns of four species. The starred spots indicate strongly anionic compounds (e.g. indoxylsulphate) which are lost on electrolytic desalting. Colours vary greatly and include browns, blues, purples, yellows, etc. The major centrally placed spot is urea. Macaque = *Macaca mulatta*; marmoset = *Hapale jacchus*; baboon = *Papio* sp. Separation by two-way chromatography using *iso*-propanol-ammonia followed by butanol-acetic-acid-water.

of the rhesus species. This surprising result was confirmed by feeding a second indole, 5-hydroxy-tryptamine, to a limited number of species in our own colony. Following administration of a limited amount of this compound (in the form of pieces of banana), man and three other species examined each excreted further indoles but these indoles were different for each different species.

DISCUSSION

The above very brief review of my results raises a number of questions, namely:

1. Can specimens obtained from different species, often living in different geographical locations, be compared? I believe they can, both because the normal human pattern remains constant over whole continents, and also because wherever it was possible to compare patterns from specimens of the same species obtained from different primate centres they gave identical patterns.

2. Are dietary differences a possible explanation of the results? This seems unlikely. In man, the only species for whom data were previously available, a very few dietary components are known to produce urinary metabolite changes. Thus, dates lead to urinary 5-hydroxy pipecolinic acid—a single substance never observed on our chromatograms—and bananas, blue cheese and some wines yield changes in indole patterns due, it is thought, to their tryptamine, hydroxytryptamine and tyramine content. It is hardly likely that wine and blue cheese form part of the staple diet in primate colonies and so precautions were taken to exclude only bananas from the diet, except in the 5-hydroxytryptamine administration experiment.

3. Could bacterial contamination be responsible for, at least, the indole patterns. Again this seems unlikely because of pattern identity from the same species derived from different colonies. Of course, the gut flora may be the source of some of the spots observed (e.g. indoxyl sulphate in man) but these spots would be considered part of the normal pattern.

4. What is the taxonomic significance of these data? It would seem that metabolic pathways of particular compounds certainly vary from species to species. This has been stressed for tryptophan and tryptamine (Smith *et al.*, 1967). Referring back to the blood amino acid patterns in Fig. 1 it will be noted that the levels of phenylalanine and tyrosine vary in different species and hence, although it was not shown here, one presumes that the levels of their urinary metabolites will vary. Other studies on so-called phenylketonuric monkeys (Boggs *et al.*, 1963), confirm this view. It would obviously be worth while to follow tracer doses of labelled single amino acids through the different species to determine the general validity or otherwise of this hypothesis.

ACKNOWLEDGMENTS

I am indebted to the various primate centres in the United States which supplied specimens for this study. The work was supported by

U.S.P.H.S. Grant DE 02232 from the National Institute for Dental Research, NIH, during the tenure of a visiting Professorship at the Institute of Dental Science, University of Texas Dental Branch, and I am particularly indebted to Dr. B. Levy for his invitation to work at that Institute.

REFERENCES

Boggs, D. H., McLay, D., Kappy, M. and Waisman, H. A. (1963). Excretion of indolyl acids in phenylketonuric monkeys. *Nature, Lond.* **200**, 76.

Efron, M. L., Young, D., Moser, H. W. and MacCready, R. A. (1964). A simple chromatographic screening test for the detection of disorders of amino acid metabolism. *New Engl. J. Med.* **270**, 1378–83.

Fooden, J. (1961). Urinary amino acids of non-human primates. *Zoologica, N.Y.* **46**, 167–80.

Gartler, S. M., Firshbein, I. L. and Dobzhansky, T. (1956). A chromatographic investigation of urinary amino acids in the great apes. *Am. J. Phys. anthrop.* **14**, 41–55.

Jepson, J. B. (1962). *In* Smith, I. (ed.) "Chromatographic and Electrophoretic Techniques", Vol. 1, 183–211. Heinemann Medical Books, London.

Smith, I. (1968). *In* Smith, I. (ed.) "Chromatographic and Electrophoretic Techniques", Vol. 2, 524. Heinemann Medical Books, London.

Smith, I. and Feinberg, J. G. (1965). "The 'Unikit' Apparatus in Paper and Thin Layer Chromatography and Electrophoresis", Shandon Sci. Co., London and Ohaus Scale Corp., N.J.

Smith, I. and Rider, L. J. (1967). Imidazole lactic acid—a component of normal human urine. *Proc. Soc. exp. Biol. Med.* **124**, 233–5.

Smith, I., Rider, L. J. and Lerner, R. P. (1967). A rapid two-way screening procedure for whole blood and urine. *Biochem. Med.* **1**, 9–14.

4. Acrylamide Gel Electrophoresis of Proteins in Plant Systematics

D. Boulter and D. A. Thurman

Department of Botany, University of Durham, and the Hartley Botanical Laboratories, University of Liverpool, England

ABSTRACT

A resumé of the methods and apparatus currently used in polyacrylamide gel electrophoresis is given. Various factors which are important in comparative studies are discussed, i.e. reproducibility of the method itself, extraction and the type of material extracted. The basis for the interpretation of band patterns is discussed both for patterns obtained using a general protein stain and for band patterns obtained using an enzyme stain. The potential usefulness of the technique is indicated by reference to the few data available at the present time and it is pointed out that data obtained by this method may often lead to misleading conclusions unless supplemented by studies on other properties of the proteins in question.

INTRODUCTION

For some years now chemical taxonomic data have slowly accumulated but, so far, no real attempt has been made to incorporate many of them into existing taxonomies. The main reason for this is that, from the taxonomic point of view, many of the data have been acquired in an unsatisfactory way. This is presumably because many of the data, initially microchemical and more recently macromolecular, have been obtained by chemists and biochemists who either did not fully appreciate the necessary rigour of taxonomic methods or had other objectives in view.

For various reasons, the real solution to the problem outlined above lies in the taxonomists themselves using chemical and biochemical methods alongside the more well-tried cytological and morphological ones. For this to happen, the taxonomist must be convinced both of the usefulness of chemical characters and of the feasibility of the methods employed. To be useful, chemical characters must be shown to be "good" taxonomic characters, i.e. they should not vary within the

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 39–48.

samples being considered, should not be susceptible to environmental modification and should show consistency, that is correlate with existing classifications constructed using other characters. For the methods to be feasible, they must be within the reach of the average taxonomist from a technical standpoint.

Of the many compounds which are found in living systems Zuckerkandl and Pauling (1965) maintain that semantides are potentially the most informative taxonomic characters of an organism. These authors classify semantides into three groups, primary (DNA molecules), secondary (m-RNA molecules) and tertiary (most proteins). Thus, in the case of proteins, apart from a certain amount of crypticity and degeneracy (see Dixon, 1966), the same amino acid sequence in the proteins of different organisms corresponds to the same sequence of base pairs in part of their DNA molecules.

The main methods used at present to compare the properties of proteins are (i) the determination of their amino acid composition, (ii) fingerprint pattern of peptides, (iii) amino acid sequence, (iv) molecular weight, (v) catalytic activity, (vi) immunological properties, (vii) electrophoretic mobility and (viii) chromatographic characteristics.

By far the most informative of these methods is the amino acid sequence. This method necessitates the isolation of a pure protein and then the determination of the amino acid sequence using methods developed by Sanger (1947) and others. It is time consuming, requires considerable knowledge of specialized techniques and some expensive equipment. At this stage in comparative studies on plant proteins it is probably only feasible with the respiratory pigment cytochrome C and even then in only certain laboratories.

The other methods vary in the degree of technical expertise required, the grade of information produced and in their suitability for surveying large numbers of taxa. In the case of disc electrophoresis, which is the subject of this paper, the method is highly feasible technically in its application to systematic surveys since it is quick, simple and inexpensive to use. The method requires a minimum of chemical knowledge and techniques. The method is presented here at a non-technical level together with some of the results obtained in systematic studies in order to evaluate the potential usefulness of the technique for such work.

METHODS AND APPARATUS OF ACRYLAMIDE ELECTROPHORESIS

The use of polyacrylamide gels for the separation of proteins as a substitute for starch gel was first reported by Davis and Ornstein (1959), and Raymond and Weintraub (1959). While developing polyacrylamide gels for protein separations, Davis and Ornstein introduced

the technique of disc electrophoresis which exploited to maximum advantage the mechanisms responsible for the separation of proteins in polyacrylamide gels. Since the original introduction of disc electrophoresis by Davis and Ornstein (1961) many variations of the method have appeared. Similarly, the technique introduced by Raymond and Weintraub (1959), using a flat bed of polyacrylamide has beeu extended and modified in various ways. The following literature references (Davis and Ornstein, 1961; Davis, 1964; Raymond, 1962) deseribe the operation of the technique in detail and need not be reiterated here.

The sample pattern produced by the separated proteins has been given a surprising number of names (Strickland, 1966) including: electrophoregram, electrophorogram, electrophoreogram, electrophoretogram, electrochromatogram, electrophoretic pattern, phoregram, pherogram, proteinogram and proteinographic chart. The term electrophoregram is now favoured (Strickland, 1966) which conforms to the parent word and sound Greek usage.

Apparatus for disc electrophoresis both on an analytical and preparative scale is now available commercially from numerous suppliers. Details for the construction of such apparatus is also available in the literature; for the construction of analytical disc electrophoresis apparatus sce (Davis and Ornstein, 1961; Laycock *et al.*, 1965; Rogers, 1965). Adaptations to a miero-scale have also been made (Pun and Lombrozzo, 1964; Grossbach, 1965; Gofman, 1965; Matioli and Niewisch, 1965) and on a preparative scale (Lewis and Clark, 1963; Racusen and Calvanico, 1964; Altschul *et al.*, 1964; Radhakrishnamurthy *et al.*, 1965; Maizel, 1964; Duesberg and Rueckert, 1965).

Apparatus previously used for vertical starch gel electrophoresis is commonly used for polyacrylamide gel electrophoresis both at the analytical and preparative levels. Raymond (1962) has described such an apparatus and this apparatus has found widespread use.

The chief advantages of disc electrophoresis over flat-bed methods using polyacrylamide gel are often quoted as (see Nerenberg, 1966) (1) unparallelled resolution of protein bands, (2) simplicity of operation and (3) minimal cost of reagents and materials; and the disadvantages as (1) inability to compare different samples on the same gel and (2) difficulties encountered in quantitation of the resolved bands. In discussion it is likely, however, that many of these points would be disputed. Thus Allen and Moore (1966), in a comparison of the flat-bed method using polyacrylamide gel with disc electrophoresis applied to human male serum protein found greater zone separation of compounds, without loss of sharpness of the separated bands, using the flat-bed method though the separation took twice as long. Again, many of the methods described for flat-bed polyacrylamide gel electrophoresis

are as simple or in some cases simpler to carry out than those used in disc electrophoresis. Some workers, carrying out comparative studies, prefer to run different samples on the same gel, and hence use a flat-bed method employing a single gel for all samples instead of a separate gel for each sample as in disc electrophoresis, though the reproducibility of the disc electrophoretic method is high (see Davis, 1964, photograph, page 419).

Disc electrophoresis is normally performed at fixed gel concentrations and in standard discontinuous buffers chosen to give narrow starting zones of the material to be electrophorized. The use of such conditions limits the choice of buffer and its pH. Hjerten *et al.* (1965) showed that similar resolution to that obtained in disc electrophoresis can he obtained in a system where a continuous buffer system is used, with only one gel concentration, provided that the starting zones are sharp enough. This was achieved simply by having the sample in a buffer of lower conductivity than that used in the electrophoresis chamber. Such systems give greater flexibility of choice of buffer types and their pH's.

Proteins separated by gel electrophoresis are usually located by using a general protein stain, or by the enzymic activity of the separated proteins, or by eluting the proteins off the gel followed by their estimation. A stain which has found widespread use is amido-black in acetic acid. This stain, however, is also taken up by the gel itself and before protein bands can be located this background stain must be removed. This is usually achieved either by washing the gel in a solvent in which the dye is soluble or by electrophoresis, which is much more rapid, though it should be realized that any unfixed protein will be lost during this procedure. Apparatus for use in the electrophoretic destaining of cylindrical gels have been described by Davis and Ornstein (1961); a rapid electrophoretic staining and destaining process for such gels is described by Schwabe (1966). Many descriptions for the electrophoretic destaining of flat-bed gels are also available (see Stanton, 1965; Burger and Wardrip, 1965). Recently Chrambach *et al.* (1967) have described a very simple and sensitive method for the location of protein bands on polyacrylamide gel cylinders by using Coomassie blue in 20% (W/v) TCA solution which does not stain the gel unless protein is present.

Many techniques have been described for the location of enzymes on polyacrylamide gels and the reader is referred to the books of Lawrence (1964) and Wilkinson (1965) which describe details for such methods.

REPRODUCIBILITY

Any technique used in comparative studies of the type described here should be highly reproducible. Band patterns obtained when samples

of the same protein extract are subjected to disc electrophoresis have been shown to fulfil this requirement, provided that the electrophoresis and staining is carried out under standard conditions. However, the method of extracting, physiological state of the material and the amount of protein subjected to electrophoresis can profoundly affect the band patterns obtained.

Extraction

The band pattern obtained with a general protein stain will vary with the type of extractant used, since different types of proteins have different solubilities in different solvents. Generally speaking, plant proteins fall into the following solubility classes: albumins, soluble in water and dilute salt solutions; globulins, soluble in salt solutions, but not or only slightly soluble in water; prolamins, soluble in 70–80% (w/v) ethanol but not soluble in absolute ethanol or in water; glutelins soluble in dilute acids and alkalis, but not in water, salt solutions or aqueous alcohol. The protein pattern obtained with any given material using these different extractants will depend upon the relative amounts and types of protein in the material extracted. However, these classes are not strictly mutually exclusive. This point is illustrated by a study on wheat grain proteins described by Barber *et al.* (1967) in which the effects of extraction with dilute acetic acid and tris/glycine buffers can be seen.

Since proteins possess positive and negative charged groups there is the possibility for protein–protein interaction to take place; the conditions which bring about such interactions are varied and complex, but important in this connection is the concentration of the protein, the nature and concentration of ions in the environment, the pH and the rH. In short, these factors must be controlled and this is achieved as far as possible by extracting under standard conditions, though it should be realized that, when extracting material from different taxa, each contributes its own different components to the conditions of the extraction procedure.

Material

Just as the extraction itself must be standardized so must be the material extracted. In practice one must extract the same organ at the same stage of development from plants grown under the same environmental conditions. It is now well established that different organs of the same plant give different band patterns (see Steward *et al.*, 1965; Barber *et al.*, 1967) and also that the same organ at different stages of development may give a different band, e.g. the protein band patterns observed in different aged leaves of *Nicotiana* (Hart and Bhatia, 1967)

and in developing seeds of *Vicia faba* (Davis, 1967). Material which has been grown under standard conditions is not always readily obtained, and the choice of organ compared may become important. Dunnill and Fowden (1965) have pointed out that seed composition is unlikely to be affected by variation in nutritional and environmental factors during growth.

INTERPRETATION

Whatever the method of locating the protein on gels after electrophoresis it is important at this stage to realize the usual assumptions made when comparing band patterns. In this method, it is not the amino acid sequence of the protein that is being compared but the protein's position on a gel after electrophoresis. This position depends on:

1. Its amino acid composition, not on the sequence in which the amino acids are arranged, and
2. A molecular sieving effect which is determined by the size and shape of the protein in question.

Thus, some of the amino acid residues of the protein will be charged, depending on the pH at which electrophoresis is carried out; in some cases, the charge is positive, in others it is negative. However, it is possible for the amino acid composition to vary considerably from one protein to another and yet for the proteins themselves to carry the same overall charges; hence if their size and shape are the same, they will *have the same electrophoretic mobility. Furthermore, proteins with* different charges may have the same electrophoretic mobility if these differences are compensated for by a size and/or shape difference. Put simply, identical gel patterns obtained with two different taxa, by no means imply that the proteins responsible for the patterns are identical.

Thus, just as with a character such as leaf shape, where the same leaf shape in different species may be due to the interaction of different protein systems, so the same protein band pattern in different species may also be due to different proteins. With closely related species, however, it is likely that similar band patterns are due to homologous proteins. The difficulty of comparing protein bands due to doubts about homology is to a large extent overcome by the use of a specific location reagent such as an enzyme stain. Recently some enzymes have been shown to exist in multiple forms in the same organism, and the number of enzymes of this type is increasing rapidly. This phenomenon must be borne in mind when comparing enzyme electrophoregrams from different species, although homologous forms can often be recognized as they can occur in characteristic amounts.

RESULTS AND DISCUSSION

From results already obtained, and here the extreme paucity of the existing data must be stressed, it appears that protein band data obtained with different taxa may prove useful at different levels.

Boulter *et al.* (1967) have examined the storage globulins of seeds of a considerable number of species from the five genera of the tribe Vicieae and have established that a tribal protein band pattern exists which was not detected in the other tribes of the family Leguminosae which they examined. These authors also suggested that tribal specific patterns might exist in other tribes of this family, although insufficient species were examined to establish this with any certainty. Fox *et al.* (1964) in a comparison of albumin proteins of a few species of the Leguminosae found that protein band patterns of species within a genus resembled one another more closely than did species belonging to different genera. Again, unpublished work from this laboratory, with various species of *Phaseolus* and *Baptisia*, indicates that when a number of species from the same genus were examined the albumin band patterns were the same, though this did not hold for some of the enzymes examined. In contrast to these findings Hart and Bhatia (1967) showed that within the genus *Nicotiana* the leaf proteins of each species exhibited a distinct and characteristic pattern, and that this was also true for various enzymes studied. Similarly, Desborough and Peloquin (1966) from a study of the protein band data obtained with proteins extracted from tubers of various species of *Solanum*, concluded that species specific patterns existed. However, when these same workers (Desborough and Peloquin, 1967) investigated esterase isoenzyme patterns in some of the same species they were unable to establish species specific patterns. Loeschcke and Stegemann (1966) from a study of the soluble proteins of *Solanum*, concluded that protein band patterns obtained were independent of the variety examined. By contrast, Hiltz and Schmitthenner (1966) in an analysis of proteins extracted from the leaves of non-resistant and resistant varieties of soybean, could find no differences between the two varieties examined. Barber *et al.* (1967) after an electrophoretic study of the glutens of three varieties of wheat found that two of the varieties, having a common parentage, possessed similar band patterns, while the third variety of different parentage was characterized by a different band pattern.

As pointed out by Boulter *et al.* (1966) there is an urgent need to establish the extent of "polymorphism" for protein types in natural populations. Although there are few data, we incline to the view that the selection of enzyme types will be more effective under natural conditions than in cultivation.

Johnson and Hall (1966) in a report on electrophoretic studies on wheat proteins state that "homology between the bands of different species, based on similarity in migration velocity, provides a criterion of genetic affinity from which evolutionary relationships may be inferred". Whilst these results and those of others described in this paper show the potential usefulness in systematics of the measurement of electrophoretic mobilities of proteins, it is our view that attempts to establish taxonomic relationships using this method will often lead to misleading conclusions unless supplemented by studies on several other properties of the proteins in question.

REFERENCES

Allen, R. C. and Moore, D. J. (1966). A vertical flat-bed discontinuous electrophoresis system in polyacrylamide gel. *Analyt. Biochem.* **16**, 457–65.

Altschul, A. M., Evans, W. J., Carney, W. B., McCourntney, E. J. and Brown, H. D. (1964). Some aspects of preparative electrophoresis on polyacrylamide gel: application to bovine serum albumin. *Life Sci.* **3**, 611–15.

Barber, J. T., Wood, H. L. and Steward, F. C. (1967). The separation of the proteins of wheat by acrylamide gel electrophoresis, with special reference to mottled and unmottled wheat. *Can. J. Bot.* **45**, 5–19.

Boulter, D., Thurman, D. A. and Derbyshire, E. (1967). A disc electrophoretic study of globulin proteins of legume seeds with reference to their systematics. *New Phytol.* **66**, 27–36.

Burger, W. C. and Wardrip, W. O. (1965). Rapid destaining of polyacrylamide gel electrophoresis slabs. *Analyt. Biochem.* **13**, 580–2.

Chrambach, A., Reisfeld, R. A., Wyckoff, M. and Zaccari, J. (1967). A procedure for rapid and sensitive staining of protein fractionated by polyacrylamide gel electrophoresis. *Analyt. Biochem.* **20**, 1–15.

Davis, B. J. (1964). Disc electrophoresis. II. Method and application to human serum proteins. *Ann. N.Y. Acad. Sci.* **121**, 404–27.

Davis, B. J. and Ornstein, J. (1959). A new high resolution electrophoresis method. Delivered at the Society for the Study of Blood at the New York Academy of Medicine.

Davis, B. J. and Ornstein, L. (1961). "Disc Electrophoresis." Preprinted by Distillation Product Industries (Division of Eastman Kodak Company) Rochester, New York.

Davis, O. J. (1967). "Studies on the Nitrogen Metabolism of Developing Seeds of *Vicia faba* (L)." Ph.D. Thesis, University of Liverpool.

Desborough, S. L. and Peloquin, S. J. (1966). Disc electrophoresis of tuber proteins from *Solanum* species and interspecific hybrids. *Phytochem.* **5**, 727–33.

Desborough, S. L. and Peloquin, S. J. (1967). Esterase isozymes from *Solanum* tubers. *Phytochem.* **6**, 989–94.

Dixon, G. H. (1966). Mechanisms of protein evolution. *In* Campbell, P. N. and Greville, G. D. (eds.), "Essays in Biochemistry", Vol. 2, 147–204. Academic Press, London and New York.

Duesberg, P. H. and Rueckert, R. R. (1965). Preparative zone electrophoresis of proteins on polyacrylamide gels in 8M urea. *Analyt. Biochem.* **11**, 342–61.

Dunnill, P. M. and Fowden, L. (1965). The amino acids of seeds of the Cucurbitaceae. *Phytochem.* **4**, 933–44.

Fox, D. J., Thurman, D. A. and Boulter, D. (1964). Studies on the proteins of seeds of the Leguminosae—1. Albumins. *Phytochem.* **3**, 417–19.

Gofman, Y. Y. (1965). Microelectrophoresis of proteins in polyacrylamide gel. *Biochemistry* (Biokhimiya) **30**, 996–1001.

Grossbach, U. (1965). Acrylamide gel electrophoresis in capillary columns. *Biochim. Biophys. Acta.* **107**, 180–82.

Hart, G. E. and Bhatia, C. R. (1965). Acrylamide gel electrophoresis of soluble leaf proteins and enzymes from *Nicotiana* species. *Can. J. Genet. Cytol.* **9**, 367–74.

Hiltz, J. W. and Schmitthenner, A. F. (1966). Electrophoretic comparisons of soybean leaf proteins from varieties resistant and susceptible to *Phytophthora megasperma* var. *sojae*. *Phytopathology* **56**, 287–91.

Hjerten, S., Jerstedt, S. and Tiselius, A. (1965). Some aspects of the use of "continuous" and "discontinuous" buffer systems in polyacrylamide gel electrophoresis. *Analyt. Biochem.* **11**, 219–23.

Johnson, B. L. and Hall, O. (1966). Electrophoretic studies of species relationships in *Triticum*. *Acta. Agric. scand. suppl.* **16**, 222–4.

Lawrence, S. H. (1964). "The Zymogram in Clinical Medicine." Thomas, Springfield, Illinois, U.S.A.

Laycock, M. V., Thurman, D. A. and Boulter, D. (1965). An improved method for the detection of dehydrogenases using tetrazolium salts. *Clinica. chim. Acta.* **11**, 98–110.

Lewis, U. J. and Clarke, M. O. (1963). Preparative methods for disc electrophoresis with special reference to the isolation of pituitary hormones. *Analyt. Biochem.* **6**, 303–15.

Loeschcke, V. and Stegemann, H. (1966). Protcine der kartoffelknollen in abhängigkeit von sorte und virosen (polyacrylamid-elektrophorese). *Phytochem.* **5**, 985–91.

Maizel, J. V. (1964). Preparative electrophoresis of proteins in acrylamide gels. *Ann. N.Y. Acad. Sci.* **121**, 382–90.

Matioli, G. T. and Niewisch, H. B. (1965). Electrophoresis of hemoglobin in single erythrocytes. *Science, N.Y.* **150**, 1824–26.

Nerenberg, S. T. (1966). "Electrophoresis." Davis, Philadelphia, U.S.A.

Pun, J. Y. and Lombrozzo, L. (1964). Microelectrophoresis of brain and pineal proteins in polyacrylamide gel. *Analyt. Biochem.* **9**, 9–20.

Racusen, D. and Calvanico, N. (1963). Preparative electrophoresis on polyacrylamide gel. *Analyt. Biochem.* **7**, 62–6.

Radhakrishnamurthy, B., Dalferes, E. R. and Berenson, G. S. (1965). A simplified column for gel electrophoresis. *Biochim. Biophys. Acta* **107**, 380–83.

Raymond, S. (1962). A convenient apparatus for vertical gel electrophoresis. *Clinica chim. Acta* **8**, 455–70.

Raymond, S. and Weintraub, L. (1959). Acrylamide gel as a supporting medium for zone electrophoresis. *Science, N.Y.* **130**, 711.

Rogers, L. J. (1965). A simple apparatus for disc electrophoresis. *Biochim. Biophys. Acta.* **94**, 324–9.

Sanger, F. (1952). The arrangement of amino acids in proteins. *Adv. Protein Chem.* **7**, 1–62.

Schwabe, C. (1966). Disc electrophoresis. *Analyt. Biochem.* **17**, 201–09.

Stanton, M. G. (1965). Rapid removal of background dye from zone electrophoretograms. *Analyt. Biochem.* **12**, 310–15.

Steward, F. C., Lyndon, R. F. and Barber, J. T. (1964). Acrylamide gel electrophoresis of soluble plant proteins: A study on pea seedlings in relation to development. *Am. J. Bot.*, **52**, 155–64.

Strickland, R. D. (1966). Electrophoresis. *Analyt. Chem.* **38**, R.99.

Wilkinson, J. H. (1965). "Isoenzymes." Spon, London.

Zuckerkandl, F. and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theoret. Biol.* **8**, 357–66.

5. The Application of Gel Electrophoresis to the Classification of Micro-organisms

J. R. Norris

"Shell" Research Limited, Milstead Laboratory of Chemical Enzymology, Sittingbourne, Kent, England

ABSTRACT

A general description of starch and polyacrylamide gel electrophoresis is given and observations on the use of these techniques for separation of microbial cell components are described. Electrophoretic analysis of proteins and various enzymes in microbial cell extracts can provide taxonomically valuable information, and taxonomic groupings based on electrophoretic patterns are compared with groupings based on more conventional characteristics.

These methods must be used and interpreted with caution and the results must be correlated with studies of other properties of the organisms concerned. Nevertheless, the results already available are encouraging and it is hoped that gel electrophoretic analysis will be extended to include many more micro-organisms and many more types of cell components.

The microbiologist is often faced with the need to identify a newly isolated organism or to prepare a detailed classification of a group. To enable him to do this he has developed a series of tests which can be applied to his material. These include such diverse topics as observations of morphology and motility, specific staining reactions, studies of the antigenic composition of cells or their products, pathogenicity, amino acid composition of cells, DNA base ratios, nucleic acid hybridization, infra-red absorption spectra, cell wall composition and a very wide range of so-called "biochemical tests" which take the form of simple questions of the type "Does this micro-organism have the ability to carry out each of a series of enzyme reactions?" By examining an organism in this way it is possible to build up a biochemical profile which may reflect many aspects of the metabolism of the organism and

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 49–56.

which may be compared with similar profiles of other organisms. Such biochemical profiles form the basis of much of our present classification of micro-organisms.

In preparing a biochemical profile the microbiologist is restricted in two ways. He can only use reactions which may be tested easily under laboratory conditions, and this precludes him from examining many aspects of microbial metabolism. Secondly, he will obtain useful information only from those reactions which are sometimes positive and sometimes uegative in the particular organisms with which he is working. Thus, for instance, if we take a range of aerobic spore-forming bacteria and test growing cultures for catalase by adding hydrogen peroxide and looking for effervescence, or for esterase by adding α-naphthyl acetate and testing for the naphthol residue by adding a diazonium salt, the answer in every case will be positive and catalase and esterase tests are assessed as reactions of no taxonomic significance in this group of bacteria.

However, enzyme molecules are complex entities, their enzymically active sites may concern only small areas of the molecule and there may be extensive scope for variation in the rest of it. What the microbiologist is studying when he determines the biochemical profiles of two organisms is the way in which the organisms cope with particular aspects of their environments. If both organisms carry out the same enzymic activity, the simple type of test usually employed gives no indication as to whether they do so by producing exactly the same enzyme molecules, i.e. that they are in this respect genetically related to each other, or whether they are producing different molecules with similar enzymic activities. The latter situation, of course, does not imply any direct relationship at the genetic level. As techniques have become available for the investigation of the molecular nature of specific enzymes it has become possible to ask more penetrating questions about the enzyme systems of micro-organisms aud the results are proving to be of considerable taxonomic interest.

Smithies (1955) showed that complex mixtures of biological materials could be separated by electrophoresis in starch gels. The pore sizes of starch gels are such that the gel matrix exercises a molecular sieving action and the rate of migration of a molecule in such a system is determined by its shape and size as well as by its net surface charge. The application of gel electrophoresis techuiques to biological materials has showu that many enzymes exist in nature as multiple-molecular species. Paul aud Fottrell (1961) showed that the serum esterases of various animals were complex when studied by starch gel electrophoresis and that the esterase patterns obtained for the different species were of taxonomic siguificance. Subsequently, other gels have been

used, particularly polyacrylamide (Ornstein, 1964) with success as matrices for electrophoresis.

The initial impetus to explore the taxonomic potential of gel electrophoresis was a practical one. *Bacillus thuringiensis* is an unusual aerobic spore-forming bacterium which causes rapidly spreading disease in insect communities. It has attracted attention for some years as a basis for microbiological insecticides. When workers at the Agricultural Research Council's Pest Infestation Laboratory at Slough began a study of an insecticide based on *B. thuringiensis* in 1960 they were surprised to find that disease was appearing in breeding stocks of caterpillars far removed from the site of their experiments. Cultures of *B. thuringiensis* were isolated from these outbreaks but it was not possible to say whether these were identical with the bacterium involved in the insecticide work or not since little was known about the classification of these insect pathogens.

If cells from a young culture of *B. thuringiensis* are harvested and disintegrated—by ultrasound for example—an extract is obtained which contains numerous proteins, many of which have enzymic activity. If such an extract is subjected to starch or polyacrylamide-gel electrophoresis the components are separated and particular enzymes may be detected by subsequently treating the gel with specific indicators. There are many suitable colour indicators for a wide range of enzymes (Pearse, 1960). Extracts of *B. thuringiensis* examined in this way revealed multiple molecular forms of catalase, peroxidase, esterase and proteinase and the resulting esterase patterns were found to be particularly valuable for taxonomic purposes. Isolates from the Pest Infestation Laboratory were of three different esterase types and Norris and Burges (1963) were able to show that the outbreaks were not due to contamination of breeding stocks with insecticide material. The source of one outbreak was traced, using esterase analysis, to an infected stock of insects introduced from a grain store in Kenya where bacterial disease was endemic.

The correlation between the patterns of esterases found in cell extracts by gel electrophoresis and the results of other approaches to the classification of the *B. thuringiensis* group are striking. Norris (1964) demonstrated a close similarity between esterase groupings and nine serotypes based on flagellar antigens. Esterase analysis has now become a generally accepted procedure in the classification of this group of bacteria (Heimpel, 1967). Application of the method to other members of the genus *Bacillus* (Norris, 1962) has served to emphasize the similarity of esterase groupings and groupings based on more conventional criteria. *Bacillus subtilis*, *B. pumilus* and *B. licheniformis* each had their own, distinctive, esterase patterns suggesting that these

species are stable and homogeneous. The patterns derived from *B. cereus* and *B. megaterium* indicated extensive heterogeneity among different strains of these species. Such results parallel closely those obtained by immunological and other biochemical criteria. Catalases, on the other hand, proved to be of little value since the catalases derived from a wide range of aerobic spore-forming bacteria had the same mobility in the starch gel electrophoresis technique.

Cann and Wilcox (1965) compared the starch gel electrophoretic patterns of catalase and esterase of a wide range of *Mycobacterium* strains with biochemical profiles for the same organisms. As with the aerobic spore-forming bacteria the molecular forms of catalase were similar in most of the bacteria studied but several forms of esterase were demonstrated in all strains examined, and characteristic patterns could be identified for each of the recognized species, *Mycobacterium phlei*, *M. smegmatis*, *M. rhodochrous* and *M. fortuitum*. *M. smegmatis* and *M. fortuitum* contained esterases which resisted heating at 100° for 10 min. The authors concluded that identification of *Mycobacterium* strains by their esterase patterns was much more rapid and at least as reliable as identification by more conventional methods.

In an investigation involving 24 strains of *Corynebacterium* Robinson (1966) applied starch gel electrophoresis to the esterases, catalases and peroxidases present in cell extracts. The human pathogens, *C. diphtheriae*, *C. ulcerans*, *C. xerosis*, *C. hofmannii* and *C. haemolyticum* were characterized by the possession of one catalase and no peroxidase. Three esterases were present in *C. haemolyticum*. The animal pathogens, *C. renale*, *C. bovis*, *C. ovis* and *C. equi*, had two catalases, one identical with that of the human pathogens. *C. pyogenes* showed no catalase or peroxidase. Plant pathogenic species contained a common peroxidase and a common catalase as well as esterases. Robinson produced evidence that esterase analysis was of value in the separation of strains or serotypes, that peroxidases could be used as indicators of plant association and that catalases may be used for the separation of taxonomic groups above the species level. Several features of interest emerged from this study. *C. poinsettiae* and *C. flaccumfaciens*, for example, gave identical catalase, peroxidase and esterase patterns thus supporting the identity of these two species proposed by Cummins (1962) on a basis of cell walls and by Lelliot (1966) on a basis of their serology. *C. manihot* and *C. aquaticum* are not plant pathogens but electrophoretic studies suggested that they were closely related to the *C. poinsettiae/flaccumfaciens* organisms. The anomalous position of *C. pyogenes* has already been noted by other workers and its place within the genus is insecure.

Lund (1965) used polyacrylamide gel electrophoresis to examine

esterases and proteins in extracts of streptococci. *Streptococcus faecalis* strains showed similar esterase patterns which differed from the pattern typical of *S. durans*. *S. faecium* failed to show a definite esterase pattern.

Williams and Bowden (personal communication) obtained similar results using starch gel separation of glucose-6-phosphate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase of streptococci. These authors also concluded that *S. faecalis* and *S. faecium* were distinct groups with *S. durans* closely related to *S. faecium*. They were tentatively able to identify three strains which did not fall into the conventional taxonomic scheme.

Cell extracts contain many different proteins, some of which have enzymic characteristics while others do not. Staining methods based on enzyme reactions are often highly sensitive and reveal the presence of an enzyme where no protein can be detected by direct staining in a starch or polyacrylamide gel. Many of the proteins present in extracts can, however, be located by staining with non-specific protein stains such as naphthalene black. Lund (1965) in her study of streptococci employed this technique using polyacrylamide for separation and naphthalene black for staining. Polyacrylamide is a better medium than starch for studies of proteins, as opposed to enzymes, since it gives better resolution of individual proteins and it is possible to detect much lower concentrations of protein in the transparent polyacrylamide than can be seen in the opalescent starch. Lund's protein results were complex but strikingly informative and showed that her streptococci fell into two groups; *S. faecalis* and its varieties, on the one hand, and the *S. faecium/durans* group on the other. Both in the protein and esterase techniques one streptococcal strain proved to be unlike all the others. It also differed in being motile and was clearly a different type of *S. faecium*.

The potential of gel electrophoresis of proteins for species identification has been demonstrated in fungi, for example, by Clare (1963) using starch for *Pythium* species, by Chang *et al.* (1962) using polyacrylamide in the genus *Neurospora* and for *Septoria* species by Durbin (1966) using polyacrylamide. Gottlieb and Hepden (1966) used polyacrylamide electrophoresis to study the proteins of various streptomycetes and concluded that the method could be used to separate or identify one streptomycete from another or a few other closely related species.

Since the work of Smithies (1955) demonstrated the potential of starch gel electrophoresis for the separation of proteins, gel electrophoretic techniques have contributed extensively to our understanding of proteins, enzymes and other large molecules. For the taxonomist

these techniques provide new means of characterizing organisms at the subcellular level. For the microbial taxonomist they have, perhaps, a special significance. Traditionally the microbiologist has asked questions about the metabolism of his organisms without enquiring too closely about the nature of the enzymes produced to bring about the reactions involved. The resulting "biochemical profiles" provide useful descriptions of micro-organisms which enable new isolates to be identified by means of character keys. However, it has long been recognized that such descriptions cannot form a basis for suggesting relationships of organisms at the genetic level. It follows, therefore, that we know little of possible evolutionary development in microbial groups. With the advent of gel electrophoresis and the demonstration that it may be applied successfully to the characterization of microbial proteins the microbiologist is now able to compare enzymes and other cell components at the molecular level and relationships so defined reflect similarities in the genetic material controlling synthesis of the molecules concerned.

The potential of this approach to taxonomic problems is well illustrated by studies in fields outside microbiology such as the impressively successful work of Vaughan and his colleagues on the seed proteins of various species of *Brassica* and *Sinapis* (Vaughan *et al.*, 1966; Vaughan and Waite, 1967a, 1967b). By a combination of serological and gel electrophoretic techniques these studies have thrown considerable light on the taxonomic status of several plants and indicated possible phylogenetic relationships.

The field opened up by these techniques is enormous. So far very few enzymes have been examined and these only in a handful of different micro-organisms. Nevertheless, the success which has already attended the application of gel electrophoresis to microbial classification is encouraging, and it is to be hoped that many more groups of micro-organisms and many more enzyme systems will soon be examined in this way. It will only be when we can build up a picture of the distribution of various types of molecules in many more microbes that we shall be able to assess the full potential of these newer analytical methods for microbial taxonomy. The least we can expect is a series of new methods for characterizing micro-organisms which will lead to the quick and accurate identification of certain isolates. It could be that, just as we learn something of man's evolutionary development by studying the tools he has produced, rather than the operations he performs with them, so we shall obtain an insight into the evolution of groups of micro-organisms by studying the molecular tools which they have developed to enable them to carry out their various metabolic processes.

REFERENCES

Cann, D. C. and Willcox, M. E. (1965). Analysis of multimolecular enzymes as an aid to the identification of certain rapidly growing mycobacteria, using starch gel electrophoresis. *J. appl. Bact.* **28**, 165–73.

Chang, L. O., Srb, A. M. and Steward, F. C. (1962). Electrophoretic separation of the soluble proteins of *Neurospora*. *Nature, Lond.* **193**, 756–9.

Clare, B. G. (1963). Starch-gel electrophoresis of proteins as an aid in identifying fungi. *Nature, Lond.* **200**, 803–4.

Cummins, C. S. (1962). Chemical composition and antigenic structure of cell walls of *Corynebacterium, Mycobacterium, Nocardia, Actinomyces* and *Arthrobacter*. *J. gen. Microbiol.* **28**, 35–50.

Durbin, R. D. (1966). Comparative gel-electrophoretic investigation of the protein patterns of *Septoria* species. *Nature, Lond.* **210**, 1186–7.

Gottlieb, D. and Hepden, P. M. (1966). The electrophoretic movement of proteins from various *Streptomyces* species as a taxonomic criterion. *J. gen. Microbiol.* **44**, 95–104.

Heimpel, A. M. (1967). A taxonomic key proposed for the species of the crystalliferous bacteria. *J. invert. Pathol.* **9**, 364–75.

Lelliot, R. A. (1966). The plant pathogenic coryneform bacteria. *J. appl. Bact.* **29**, 114–18.

Lund, B. M. (1965). A comparison by the use of gel electrophoresis of soluble protein components and esterase enzymes of some group D streptococci. *J. gen. Microbiol.* **40**, 413–19.

Norris, J. R. (1962). Electrophoretic analysis of bacterial esterase systems—an aid to taxonomy. *J. gen. Microbiol.* **28**, vii.

Norris, J. R. (1964). Classification of *Bacillus thuringiensis*. *J. appl. Bact.* **27**, 439–47.

Norris, J. R. and Burges, H. D. (1963). Esterases of crystalliferous bacteria pathogenic for insects: Epizootiological applications. *J. insect Path.* **5**, 460–72.

Ornstein, L. (1964). Disc electrophoresis. 1. Background and theory. *Ann. N.Y. Acad. Sci.* **121**, 321–49.

Paul, J. and Fottrell, P. F. (1961). Tissue-specific and species-specific esterases. *Biochem. J.* **78**, 418–24.

Pearse, A. G. E. (1960). "Histochemistry, Theoretical and Applied", 2nd Ed. Churchill, London.

Robinson, K. (1966). An examination of *Corynebacterium* spp. by gel electrophoresis. *J. appl. Bact.* **29**, 179–84.

Smithies, O. (1955). Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. *Biochem. J.* **61**, 629–41.

Vaughan, J. G., Waite, A., Boulter, D. and Waiters, S. (1966). Comparative studies of the seed proteins of *Brassica campestris, Brassica oleracea* and *Brassica nigra*. *J. exp. Bot.* **17**, 332–43.

Vaughan, J. G. and Waite, A. (1967a). Comparative studies of the seed proteins of certain species of *Brassica* and *Sinapis*. *J. exp. Bot.* **18**, 100–9.

Vaughan, J. G. and Waite, A. (1967b). Comparative electrophoretic studies of the seed proteins of certain amphidiploid species of *Brassica*. *J. exp. Bot.* **18**, 269–76.

6. Species Differences in the Amino Acid Sequences of Bacterial Proteins

R. P. Ambler

Department of Molecular Biology, University of Edinburgh, Scotland

ABSTRACT

The feasibility of comparing the chemical structure of proteins by amino acid analysis of purified preparations, by peptide maps, and by amino acid sequence determination is discussed. Comparative results obtained by these methods are described for a number of bacterial protein systems, including azurins, penicillinases and flagellins.

The study of the amino acid sequences of bacterial proteins is likely to be a useful aid in taxonomy. If two or more species contain proteins with apparently homologous sequences, an evolutionary connection between the species is suggested. The complete determination of the amino acid sequence of a protein still involves considerable effort, but the difficulties are commonly exaggerated, and methods currently available are rapid enough to enable taxonomically interesting data to be accumulated.

Bacteria contain some enzymes with activities analogous to those of enzymes in higher organisms, but in no case is it yet accepted that the bacterial protein and the plant or animal protein are homologous. The similarities between *Pseudomonas* cytochrome c-551 (Ambler, 1963) and higher cytochromes c are so slight (Sackin and Sneath, 1965) that an independent origin is credible, and there seems to be no homology between the pancreatic and the *Bacillus subtilis* DFP-sensitive proteases (Smith, 1967). The sequences of two bacterial and two higher plant ferredoxins have been determined; there are some similarities between the spinach and the clostridial protein (Matsubara *et al.*, 1967), but the latest report (Margoliash *et al.*, 1967) is that the sequences at the active sites of the plant and the bacterial proteins do not conform.

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy," edited by J. G. Hawkes, 1968, pp. 57–64.

There is little published information about how corresponding proteins from different bacterial species, orders or genera differ from each other. A selection of antigenically different *Salmonella* flagellins have been examined by amino acid analysis and peptide mapping (Ambler, 1960; McDonough, 1965; see below and Table 1). Smith (1967) has

TABLE 1. Amino acid composition differences between related proteins

	% Difference	
	Minimum (from amino acid composition)	Actual (from sequence)
Replicate analysis of identical samples	1–3	
Flagellins		
antigenically identical	1–4	
antigenically related but distinct	3–5	
phases of natural diphasic strain	7	
most different *Salmonella* types	15–20	
Salmonella/Proteus	15–20	
Azurins (between genera)	9–13	32–39
Penicillinase		
Bacillus licheniformis strains	2	1–2
B. licheniformis/S. aureus	16	? no homology
Penicillinase/flagellin	14	

shown that the DFP-sensitive proteases from two strains of *B. subtilis* (Carlsberg subtilisin and Nagarse) differ in 84 out of 275 amino acid positions; the remaining 191 matching amino acids are distributed throughout the molecule.

There are several levels at which comparisons of proteins from different organisms can usefully be made. These are:

(a) physical or enzymic properties (e.g. electrophoretic mobility or enzyme kinetic parameters) of the intact protein molecules, which need not necessarily have been purified;

(b) amino acid analyses of the pure proteins;

(c) peptide maps;

(d) partial or complete amino acid sequence determination.

Investigations at the first level are dealt with in other papers at this symposium (e.g. Norris, *this symposium*, p. 49). Comparisons based on amino acid composition differences are difficult to interpret, and reliable data are difficult to obtain. The results may be completely invalid if each protein is not completely pure, and meticulous care is necessary to attain sufficient accuracy to deduce correctly the number of each type

of amino acid residue in even a small protein. The results shown in Table 1 demonstrate that the differences in amino acid composition between proteins that differ in less than two-fifths of their amino acid positions are likely to be nearly as great as the composition differences between randomly selected proteins. The accuracy of amino acid analysis is insufficient to distinguish between proteins which differ in only 1–2% of their amino acid positions. Peptide maps are satisfactory for comparing very similar proteins or for detecting the results of some one-step mutations, but are useless for showing similarity between proteins that have many differences.

The technical difficulties in investigating the chemical structure of proteins are sufficiently great to limit the choice of systems for comparative studies. The proteins to be investigated should be as small as possible, since although present methods are certainly adequate for solving the structure of proteins of molecular weight up to about 40 000, much faster progress can be made with smaller molecules. The proteins must be available in adequate amounts (hundreds rather than tens of milligrams), and the purification methods should be so simple that preparations from a number of different organisms can be performed in parallel.

In Table 2 the properties of some bacterial protein systems are summarized. The investigations have been continued over several years, and during this period the techniques of protein chemistry have developed very greatly, and the difficulties that caused systems to be abandoned may now be soluble.

Salmonella FLAGELLINS

The flagella can be removed from motile *Salmonella* by shaking, and purified by differential centrifugation. The organelles are disaggregated by dilute acid or urea to form a soluble protein, flagellin, which probably has a molecular weight of about 40 000 (the evidence is not quite unequivocal). At the time of investigation the size of the molecule was considered too great for complete sequence determination to be attempted, and so the results obtained were derived from amino acid compositions and peptide maps (Ambler, 1960; McDonough, 1965). Antigenically related proteins were found to be similar in both composition and peptide maps, whilst some flagellins (which do not cross-react immunologically) showed little or no peptide map resemblance, and differed in composition as much as randomly selected proteins would. Experiments of this type were incapable of providing any evidence for homology between the two types of flagellin produced by diphasic *Salmonella*.

TABLE 2. Bacterial proteins examined for species differences

Protein	Source	Molecular weight	Amino acid analysis	Peptide map	Amino acid sequence
Flagellin	*Salmonella* (many serotypes)	?40 000	*a*	*a*	—
	Proteus vulgaris	17 000	*a*	—	—
Cytochrome c	*Pseudomonas fluorescens* c-551	9 000	*a*	*a*	*a*
	c-556	?11 000	*a*	*a*	(part)
	c-552	?11 000	*a*	*a*	—
Penicillinase	*Staphylococcus aureus*	30 000	*a*	*a*	(*a*)
	Bacillus licheniformis	30 000	*a*	*a*	(part)
	Escherichia coli	?17 000	*a*	—	—
	Aerobacter/Klebsiella	?17 000	*a*	—	—
Azurin	*Ps. fluorescens*	14 000	*a*	*a*	*a*
	Bordetella bronchiseptica		*a*	*a*	*a*
	Alcaligenes denitrificans		*a*	*a*	(*a*)
	A. faecalis		*a*	*a*	(*a*)
	"*Ps. denitrificans*"		*a*	*a*	—
	Agrobacterium tumefaciens		*a*	*a*	—

[a] Levels at which proteins have been investigated.

Pseudomonas c-TYPE CYTOCHROMES

Cytochrome c, recognized and defined by a highly characteristic absorption spectrum, occurs in higher organisms, and widely though discontinuously in bacteria. The complete amino acid sequence of one bacterial protein, *Pseudomonas fluorescens* cytochrome c-551 has been determined (Ambler, 1963). In this one organism there are in addition at least two other c-type cytochromes, which all differ in isoelectric points, terminal residues, amino acid composition, and sequence around the haem group (Ambler, unpublished results). It seems that there is neither homology between the cytochromes in the one organism, nor between any of these cytochromes and the proteins from higher organisms.

PENICILLINASES

Penicillinases have been purified from *Staphylococcus aureus, Bacillus lichenformis, Bacillus cereus, Escherichia coli* and *Klebsiella* spp. The enzymes from the first two organisms have very similar molecular weights and enzymic properties (Citri and Pollock, 1966). There are at least two antigenically distinct types of staphylococcal protein (Richmond, 1965), but the chemical differences between them seem to be small. The amino acid sequence of one type has been determined (Ambler, to be published), and the amino acid alterations in some mutenes located. Natural isolates of *B. licheniformis* produce one of two enzymically distinct penicillinase types, but present evidence is that these enzyme types differ in only a very small number of amino acid residues. There is no evidence at present for any homology at all between the *S. aureus* and the *B. licheniformis* proteins (R. J. Meadway, unpublished work).

AZURINS

Azurins are a class of blue copper-containing proteins with a function in electron transport that has not yet been precisely defined. The protein was first found in *Pseudomonas* (Verhoeven and Takeda, 1956; Horio, 1958), but has since been found to occur in several other genera, including *Bordetella, Alcaligenes* and *Agrobacterium* (Sutherland and Wilkinson, 1963; Sutherland, 1965). The different azurins have quite widely different isoelectric points and amino acid compositions, but similar molecular weights. When the amino acid sequence of *Pseudomonas fluorescens* azurin had been determined (Ambler and Brown, 1967), peptide maps of other azurins were examined, but no similarities

	P	B	D	F
1	Ala	Ala	Ala	
	Glu	Glu	Glu	Ala
	Cys	–	–	–
	Ser	Ser	Glu	Asp
	Val	Val	Ala	Val
	Asp	Asp	Thr	Ser
	Ile	–	–	–
	Glu	Ala	Glu	Glu
	Gly	Gly	Ser	Gly
10	Asn	Thr	Asn	Asn
	Asp	–	–	–
	Gln	Gln	Ala	Ser
	Met	–	–	–
	Gln	–	–	–
	Phe	Phe	Tyr	Phe
	Asn	Asp	Asp	Asn
	Thr	Lys	Leu	Thr
	Asn	*Lys*	*Lys*	*Lys*
	Ala	Ala	Glu	Ser
20	Ile	Ile	Met	Ile
	Thr	Glu	Val	Val
	Val	–	–	–
	Asp	Ser	Asp	Asp
	Lys	–	–	–
	Ser	Ser	Ser	Thr
	Cys	–	–	–
	Lys	–	–	–
	Gln	Gln	Gln	Glu
	Phe	–	–	–
30	Thr	–	–	–
	Val	Val	Val	Ile
	Asn	Asn	His	Asn
	Leu	–	–	–
	Ser	Lys	Lys	Lys
	His	–	–	–
	Pro	Thr	Val	Thr
	Gly	–	–	–
	Asn	Lys	Lys	Lys
	Leu	Leu	Met	Leu
40	Pro	Pro	Ala	Pro
	Lys	Arg	Lys	Lys
	Asn	Asn	Ala	Ala
	Val	Val	Val	Ala
	Met	–	–	–
	Gly	–	–	–
	His	–	–	–
	Asn	Asn	Asn	Ser
	Trp	Trp	Trp	Val
	Val	–	–	–
50	Leu	Leu	Leu	Val
	Ser	Thr	Thr	Asp
	Thr	Lys	Lys	Lys
	Ala	Thr	Glu	Lys
	Ala	Ala	Ala	Ser
	Asp	–	–	–
	Met	Met	Lys	Glu
	Gln	Gln	Glu	Ser
	Gly	Ala	Gly	Ala
	Val	–	–	–
60	Val	Glu	Ala	Ala
	Thr	Lys	Thr	Thr
	Asp	–	–	–
	Gly	–	–	–
	Met	Ile	Met	Met
	Ala	Ala	Asn	Lys

	P	B	D	F
66	Ser	Ala	Ala	Ala
	Gly	–	–	–
	Leu	–	–	–
	Asp	Asp	Ala	Asn
70	Lys	Asn	Gln	Asn
	Asp	Gln	Asp	Asp
	Tyr	–	–	–
	Leu	Leu	Val	Val
	Lys	–	–	–
	Pro	Ala	Ala	Ala
	Asp	Gly	Gly	Gly
	Asp	–	–	–
	Ser	Thr	Thr	Glu
	Arg	–	–	–
80	Val	–	–	–
	Ile	Leu	Ile	Ile
	Ala	–	–	–
	His	–	–	–
	Thr	–	–	–
	Lys	Lys	Lys	Ser
	Leu	Val	Val	Val
	Ile	Leu	Ile	Ile
	Gly	–	–	–
	Ser	Gly	Gly	Gly
90	Gly	–	–	–
	Glu	–	–	–
	Lys	Ser	Ser	Thr
	Asp	–	–	–
	Ser	–	–	–
	Val	–	–	–
	Thr	–	–	–
	Phe	–	–	–
	Asp	–	–	–
	Val	–	–	–
100	Ser	Ala	Ser	Ser
	Lys	–	–	–
	Leu	–	–	–
	Lys	Ala	Thr	Lys
	Glu	Ala	Pro	Glu
	Gly	–	–	–
	Glu	Asp	Glu	Glu
	Gln	Asp	Ala	Asp
	Tyr	–	–	–
	Met	Thr	Ala	Ala
110	Phe	Phe	Tyr	Phe
	Phe	–	–	–
	Cys	–	–	–
	Thr	Ser	?	Ser
	Phe	Phe	?	Phe
	Pro	Pro	?	Pro
	Gly	Gly	?	Gly
	His	His	?	His
	Ser	Gly	?	Trp
	Ala	Ala	?	Ser
120	Leu	Leu	?	Ile
	Met	–	–	–
	Lys	–	–	–
	Gly	Gly	Gly	?
	Thr	Thr	Thr	?
	Leu	Leu	Leu	?
	Thr	Lys	Lys	Lys
	Leu	–	–	–
	Lys	Val	Ser	Gly
		Asp	Asn	Ser
		(C-terminus)		

P = *Pseudomonas fluorescens* B = *Bordetella bronchiseptica*
D = *Alcaligenes denitrificans* F = *Alcaligenes faecalis*

FIG. 1. Amino acid sequences of bacterial azurins. A broken line indicates that the residue is the same in all four proteins.

were apparent. The amino acid sequence of three further azurins have since been investigated (Fig. 1), and it is now clear that about half of the sequence is common to all four proteins. Each appears to differ from each of the others in about one-third of the amino acid positions, and the similarities and differences are distributed throughout the polypeptide chain. The differences in overall amino acid composition are almost as large as between randomly selected proteins (Table 1).

CONCLUSION

Peptide maps or amino acid analysis are able to indicate the differences and similarities between proteins that are nearly identical. When the amount of difference is greater, as between corresponding proteins of different bacterial genera, it may be necessary to determine the complete amino acid sequence before any chemical similarity can be detected. Present-day methods are sufficiently rapid to enable the complete structures of a series of proteins from different organisms to be determined, and such results will soon be available for taxonomic assessment.

REFERENCES

Ambler, R. P. (1960). Structural studies on bacterial proteins. Ph.D. Thesis, University of Cambridge.

Ambler, R. P. (1963). The amino acid sequence of *Pseudomonas* cytochrome *C*-551. *Biochem. J.* **89**, 349–78.

Ambler, R. P. and Brown, L. H. (1967). The amino acid sequence of *Pseudomonas fluorescens* azurin. *Biochem. J.* **104**, 784–825.

Citri, N. and Pollock, M. R. (1966). The biochemistry and function of β-lactamase (penicillinase). *Adv. Enzymol.* **28**, 237–323.

Horio, T. (1958). Terminal oxidation systems in bacteria. I. Purification of cytochromes from *Pseudomonas aeruginosa*. *J. Biochem., Tokyo* **45**, 195–205.

McDonough, M. W. (1965). Amino acid composition of antigenically distinct *Salmonella* flagellar proteins. *J. molec. Biol.* **12**, 342–55.

Margoliash, E., Perini, F. and Keresztes-Nagy, S. (1967). Structure of *Alfalfa* ferredoxin. *Proc. 7th Int. Congr. Biochem., Tokyo,* General Session A.68, p. 595.

Matsubara, H., Sasaki, R. M. and Chain, R. K. (1967). The amino acid sequence of spinach ferredoxin. *Proc. natn. Acad. Sci., U.S.A.* **57**, 439–45.

Richmond, M. H. (1965). Wild-type variants of exopenicillinase from *Staphylococcus aureus*. *Biochem. J.* **94**, 584–93.

Sackin, M. J. and Sneath, P. H. A. (1965). Amino acid sequences in proteins: a computer study. *Biochem. J.* **96**, 70P–71P.

Smith, E. L. (1967). Amino acid sequences and some properties of subtilisins. *Proc. 7th Int. Cong. Biochem., Tokyo,* Symposium I.1, 5, p. 9.

Sutherland, I. W. (1966). The production of azurin and similar proteins. *Ark. Mikrobiol.* **54**, 350–7.

Sutherland, I. W. and Wilkinson, J. F. (1963). Azurin, a copper protein found in *Bordetella*. *J. gen. Microbiol.* **30**, 105–12.

Verhoeven, W. and Takeda, Y. (1956). The participation of cytochrome *c* in nitrate reduction. *In* McElroy, W. D. and Glass, W. B. (eds.) "Inorganic Nitrogen Metabolism", pp. 159–62. Johns Hopkins Press, Baltimore.

Note added in proof: The amino acid sequence of *Agrobacterium tumefaciens* azurin is very similar to that of *Alcaligenes faecalis* azurin, although there is at least one difference, Asn (not Ser) at position 47.

SECTION II

LARGE MOLECULES (SEROLOGY)

7. Serotaxonomy of Vertebrate Soluble Proteins

P. G. H. GELL

Department of Experimental Pathology,
University of Birmingham, England

ABSTRACT

Serotaxonomy is considered under two aspects: as applying to the evolution of whole organisms, and to that of single functional types of molecule. For the first type of investigation, immuno-electrophoresis is especially applicable, combined with absorption methods. For the second, in addition to these gel tests, enzymatic digestions with analysis of fragments and ultimately full sequence analysis with evaluation of homological runs of amino acids are necessary. Examples of both these kinds of research are given.

In this discussion I hope that it will be understood that I write as an immunologist. Although I have not been able to resist some taxonomic speculations I am aware that I am an amateur, indeed an ignoramus, in this field; I hope that any naïveté here will be treated with sympathetic toleration.

The field I want to consider is a large one, and I have to ignore much early work which would be included in a full-scale review, to describe only a few illustrative examples from recent studies. In my title I have confined myself to soluble proteins, so I omit a very large body of work on cells, particularly red cells and leucocytes.

The basic methods of such work are, first, the immunization of animals with antigens to raise antibodies and then the analysis of the reactions of the antibodies so raised with preparations of related antigens. This can be done by quantitative precipitation methods in solution, but the newer methods of gel-precipitation are so much simpler and give so much more information that I shall deal only with these. If antigen A (Fig. 1) aud antibody to it diffuse towards each other from holes in agar, a line of precipitation is formed between them; if in an

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 67–76.

adjacent hole antigen, A′, related to A is allowed to diffuse it also will form a contingent precipitation line; but since some of the antibodies will react only with A and not with A′, a *spur* will be formed on the line of A whose intensity gives a measure of the difference. If the antiserum is saturated with A′, so that no further interaction occurs with it, a serum will result strictly specific to A. This is an absorption test (Fig. 1).

Where we have multiple antigen–antibody systems, as when an antiserum is raised against a crude plant extract or whole serum, that is, against a mixture of many different sorts of protein, this is conveniently displayed by immuno-electrophoresis, where a preliminary separation in agar of the proteins of the mixture according to charge is performed under the influence of an electric current and the antiserum

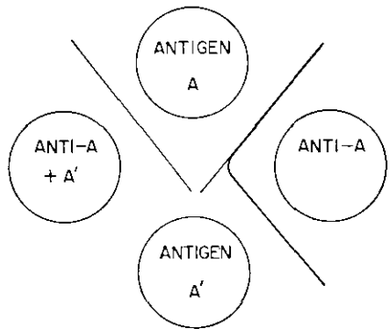


FIG. 1. Spur formation and absorption.

is put in a long trough instead of a hole. Comparison of the patterns formed by two antigen mixtures against an antiserum against one of them gives information as to the presence of a whole range of reciprocally similar proteins, though no information as to whether they are actually identical is given by this technique. Absorption of the serum can however be done with A′, and the presence or absence of reactivity with another antigen mixture, say B, can give a measure of the relative similarity of A′ and B to A. In fact we can never tell by immunological methods whether two antigens are *entirely* identical, only that they are more or less similar. I must emphasize that a pure protein is composed of many, ten or a hundred, potentially antigenic sites, that is, sites which react with and can stimulate antibodies. These sites, which we call determinants, are wholly different from one another chemically, and are of the order of molecular weight of 500–1 000, whether they are composed of amino acids of a peptide chain, of carbohydrate groups, or of attached chemical determinants. Thus the unit system we are dealing with is really an antibody–determinant reaction, rather than an

antibody–antigen reaction. What determinants actually do stimulate antibodies depends a little on chance individual differences, but mainly on the immunization situation. When a rabbit is immunized with say human gamma-globulin (IgG), sites on one part of the molecule, the Fc piece (Fig. 2), are most powerful in stimulating antibodies; when a rabbit is immunized with the gamma-globulin of another rabbit, these particular sites do not function, being identical on the two individuals; instead, sites on another part of the molecule, the Fab piece, stimulate antibodies.

We can thus set our point of discrimination at any required level, according to which question we wish to answer. For example, rabbit

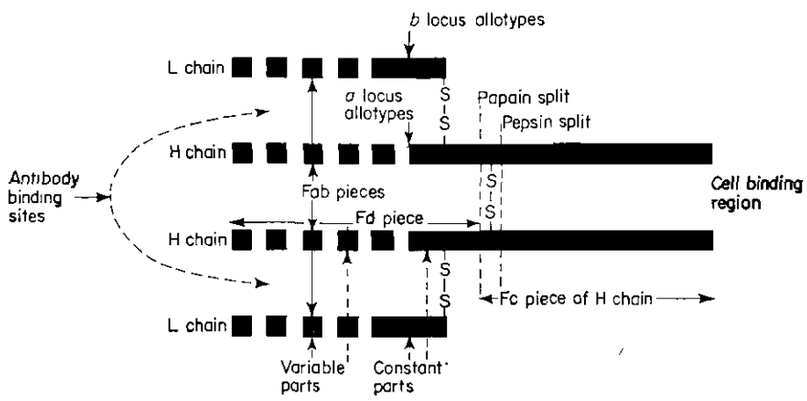


FIG. 2. Diagram of IgG molecule: modification of model suggested by R. R. Porter. (From Kelus and Gell, 1967.)

antiserum can tell quite easily the difference between human and bovine IgG; it cannot usually tell the difference between human and chimpanzee IgG. To pick out this difference we have to immunize a chimpanzee with human IgG, or *vice versa*. To pick out with the greatest precision differences between individual humans' proteins, dependent on genetic differences, we have to cross-immunize humans. Antibodies appearing by chance as a result of multiple blood transfusions have been shown (Allison and Blumberg, 1961) to discriminate between genetic markers on human lipoproteins, for example. A similar situation can be easily demonstrated for immunoglobulins in rabbits and other experimental animals, whom we can immunize much more intensively than we can men; these are the so-called allotypes, which I shall discuss below.

For purposes of serotaxonomy we can take tissues, such as the seeds or tubers of plants, or the sera or eye lenses of animals, make soluble

preparations, and immunize rabbits with them. We shall find with such substances evidence of the presence of many antigens; the number of antigens shared between species can be identified and counted. We may find that certain antigens or groups of antigens appear or disappear as the species evolve from the more to the less primitive, and we can then correlate these changes with the morphologist's hypotheses as to the inter-relations of those species. This is the straightforward approach in which we ignore and discount as far as we can small differences between the individual antigens. Intermediate forms in which a protein of one species is very slightly like the analogous protein of another, so that only some antisera may give very weak cross-reactions, are more a nuisance than anything else.

On the whole it is wise to keep this essentially pragmatic approach separate from the approach in which we consider each functional molecule as a separately evolving system, although we can sometimes analyse the same experimental material in the two opposed ways. Evolution does of course actually happen by means of the accumulation of small changes of this sort, by the alteration of DNA base triplets and their corresponding amino acids. But the way the zoologist regards evolution, as speciation, seems almost a different science from that studied by the biochemist, ultimately on the basis of the total sequence analysis of proteins, which one may call molecular evolution. The evolution of particular kinds of molecules in this sense may be quite out of step, either faster or slower, than the morphological evolution of the species as a whole; while some molecules, such as insulin or cytochrome, may change slowly, others such as the immunoglobulins may change rather fast. I suppose this is correlated with the tightness of the specification of various sorts of molecule, whether or not it has to keep its pattern relatively uniform in order to perform its function efficiently.

In summary, therefore, in tackling a taxonomic problem by serology we need to keep three things in mind: (i) that there are many determinants on antigens which may or may not "register" according to our immunization set-up; (ii) that some kinds of molecules or tissues will *a priori* be likely to show greater alterations during evolution than others; (iii) that in practice we can concentrate either on all-or-none changes in mixtures of many antigens (in which case immuno-electrophoresis is the method of choice), or on single kinds of molecule, preferably pure and of known function, in which case a careful study of spur-formation, absorption, and ultimately chemical analysis will be of most use.

What seems to me, as an immunologist, to be an extremely elegant series of studies using the first approach was published by Manski, Halbert and co-workers in the *International Archives of Allergy* (1967).

This is concerned with the proteins of the lens of the eye, which, though multiple, seem to have evolved together and comparatively slowly, presumably because they serve essentially an identical purpose in all eyed animals; or to be precise, certain groups of antigens persist apparently unchanged while new groups appear at each stage of phylogeny. In the paper I am quoting rabbits were immuuized with lens extracts from a number of marine animals, ranging from the lamprey to the shark and carp. By immuno-electrophoretic analysis of the antisera against lens extracts from a wider group of animals including mammals

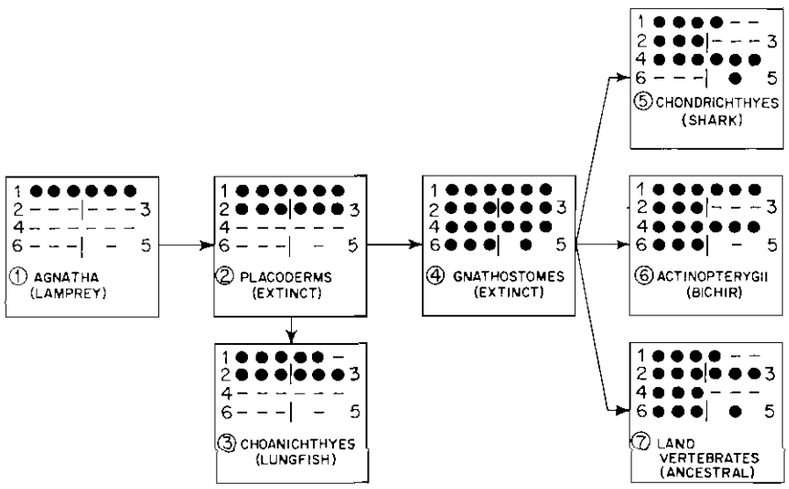


FIG. 3. Phylogeny based on lens antigens (adapted from Manski *et al.*, 1967).

and birds, and elaboratc absorption procedures into which I cannot go, they identified the various groups of antigens appearing at each level of evolution; from these the composition of the lenses of extinct groups could be deduced and a phylogenetic tree constructed. Their results are summarized in Fig. 3, adapted from their recent paper. It will be seen that some of the antigens of the lamprey persist throughout, while certain groups of antigens appeariug in the extinct placoderms and gnathostomes are deducible from those present in the families which are presumably derived from them. My own knowledge of piscine taxonomy is far too sketchy for mc to be able to judge whether such conclusions are acceptable, or even helpful, to zoologists or not. But these studies do seem to me to be a clear demonstration of the way to use serotaxonomy at this level of discrimination.

Another example is from a paper published by Kirsch and Poole (1967). They immunized grey kangaroos living in eastern Australia with serum of those living in western Australia, and *vice versa*. About three serum antigens were found to be uniformly mutually immunogenic; though mixed populations did not interbreed, sera from hybrids bred in captivity reacted with antisera against both groups. This is a pretty example because it illustrates a case intermediate between an allotype and a biochemical species. One might easily find by chance among rabbits, say, groups in which one allotype (that is a Mendelian determinant on analogous proteins, see below) was mutually lacking in almost all members of the group. But, assuming that there is no linkage in the genes coding for these different proteins, the fact that several are uniformly lacking is evidence for the evolution of a geographical species at a very early stage.

I do not wish to quote more examples of this sort of approach although there may be other and better ones of which I am ignorant. I should like to consider now the immunoglobulin molecule, in which as an immunologist I have of course a proprietary interest, and start off with an example from our own laboratory, which is taxonomically trivial but illustrates the relevance of considering the part of the molecule which is acting as immunogen. There exists in man an autoantibody, occurring in rheumatoid arthritis, and known as Rheumatoid Factor. This reacts with slightly damaged human IgG, the damage having occurred in the Fc part of the molecule (Fig. 2), where, as I have mentioned earlier, the "species" determinants lie. It will also react to a greater or lesser extent with IgG from a number of other species damaged in the same way (Fig. 4) (Stanworth and Henney, 1964)—that is, by moderate heating. Here we have a new determinant revealed by the damage, common in some degree to several mammalian species. Although the species producing the antibody is that synthesizing the antigen, the determinant recognized is to all intents and purposes equally "foreign" to all species. It will be seen from Fig. 4, which includes data also for a rabbit anti-human gamma antibody, that this is consistent with the experimental results, since discrimination between species is not very different for the so-called "auto-immune" human antibody as compared with the hetero-immune rabbit antibody. This sort of result has its significance for taxonomy, but for the taxonomy of one part of a single kind of molecule, not for the organism as a whole. It is at this point that the taxonomy and evolution of molecules begin to contrast sharply with the taxonomy and evolution of organisms.

I want now to diverge to a field which, though sufficiently related to the taxonomy of species to be I hope relevant, is more properly that of the phylogeny of functions, such as muscular contraction or, in this

case, immunological reactivity. For this I will try to summarize a great deal of still speculative and preliminary work on the evolution of the immunoglobulins. In the mammals, immunoglobulins are mainly IgG, the ordinary gamma-globulin comprising most antibodies, with a sedimentation constant of 7, and four chains in identical pairs as in Fig. 2. There is also IgM, sedimentation constant 19, with five or six pairs of chains; the light chains are similar or identical to those of IgG but the heavy chains are very different. There is also IgA, basically with the

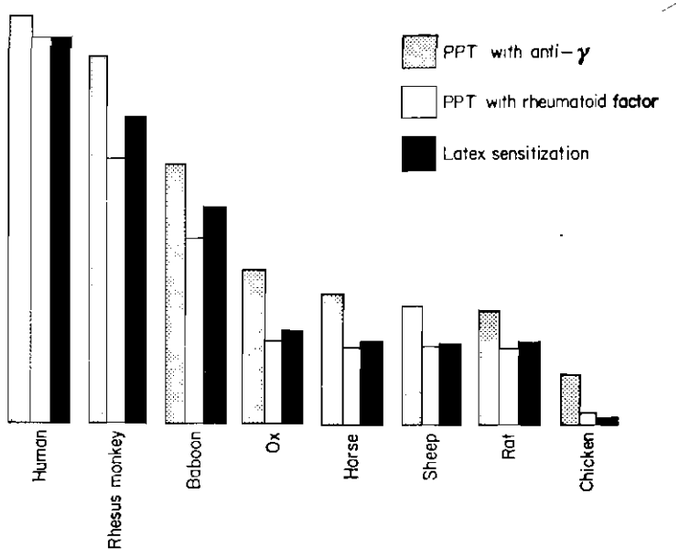


FIG. 4. Comparative amounts of precipitate produced by heat-damaged IgG of various species with a rabbit anti-human IgG antiserum and a human "rheumatoid factor" serum (also latex-fixation test with the latter serum). (From Stanworth and Henney, 1964.)

chain structure of IgG but again a characteristic heavy chain; it tends to associate in pairs of molecules to make a 10S complex. There are at least one and possibly more other types also present in low concentration.

The lowest organism in which immunological reactivity can be identified is the lamprey. In this such antibodies as can be produced appear in a fast γ or a slow β fraction of about 9S; perhaps something like an IgA. In higher fishes antibody appears associated with macroglobulins of various sizes; in amphibia the pattern approaches that of

birds and mammals. This lightning survey is based on the paper by Pollara *et al.* (1966), in a recent book "The Phylogeny of Immunity", which contains a great deal of fascinating new data which I have to ignore for lack of space.

Now current hypotheses on the evolution of the immunoglobulin molecule are based on the structural, chemical and immunological analysis of still a limited number of species, nearly all mammals, because even the identification of immunoglobulins in key species such as the lamprey is, as I have indicated, only tentative. However, one theory is put forward by Hill *et al.* (1966). Their assumption is that the basic unit coded for a 104-residue chain. This was duplicated to form the light chains, and quadruplicated to form the heavy chains, of all types of immunoglobulins.

The evidence for such events is still tenuous, but there are a few facts which make it reasonable. First of all, it seems clear that about half the light chains, and rather more than half the heavy chains, at the C-terminal ends, are fairly constant in composition; but the other bits are very variable, not merely between species, and between individuals of a species but also between molecules in a single individual. It is this variability at the N-terminal ends which makes possible the heterogeneity of binding sites needed for antibody formation. If however one compares amino acid sequences in these constant bits, there are significant runs of identically ordered amino acids, not merely when one compares Fc pieces or L chains of species as different as man and mouse, but also even when one compares light with heavy chains. Some homologies also appear if one compares the two halves of the heavy chain or of the Fc piece with one another. Many of the most recent data on this were presented at the Cold Spring Harbor Symposium in 1967, vol. XXXII, and is too complex for me to quote.

Not much of this extremely laborious sequence work has been done on immunoglobulins other than IgG. The evidence that the heavy chains of other immunoglobulins were involved in the same evolutionary process comes from work on the allotypes. Here we return to studies rather closer to serotaxonomy, and I should like to outline current studies in this field before completing the argument.

Back in 1902 Schutze cross-immunized rabbits with each other's serum and obtained precipitins, but it was not until 1948 that Oudin (1956) at the Pasteur Institute demonstrated clearly that the purified gamma globulins of one rabbit injected into others were often immunogenic. Much work by him, and by Dubiski, Kelus, Dray and others over the last fifteen years, has been devoted to this phenomenon, which Oudin named Allotypy, and the determinants, Allotypes (reviewed by Kelus and Gell, 1967). It is now known that the rabbit has two gene loci

determining allotypic determinants, one for the light and one for the heavy chain. There are three known alleles for the heavy chain and probably four for the light chain; these genes are not linked and are inherited in straight Mendelian fashion. The allelic products can be detected by antisera raised in rabbits lacking that determinant, by ordinary precipitation methods. (I may mention in parenthesis that we are in process of analysing the serum of hares for determinants related to these allotypes, but the system is highly complex; there is however a species determinant in hare IgG which can immunize rabbits, and a number of other serum proteins are also immunogenic.) To return to the evolution of immunoglobulins—it seems almost certain that the allotypic determinants of IgG molecules are all present on IgM and IgA molecules as well. The light chains of all types are identical, and under the control of a single gene—the chains are put together in the cell cytoplasm, not evidently on the ribosome—and so for the light chain allotypes this is not surprising. But for the various sorts of heavy chains this is paradoxical since they are immunologically and chemically different from one another, and clearly under the control of quite different genes. Moreover the IgM of the rabbit has itself another system of allotypic determinants in addition, quite different from the IgG determinants and not represented on IgG molecules (Kelus, in preparation). The same may well be true of IgA.

We are ignorant as yet about the location on the IgM and IgA chains of their allotypic determinants, and indeed of the class specific determinants too, and without this knowledge it is profitless to speculate as to how the suggested gene duplications fit together. We can however deduce that the allotypic determinants characteristic of IgG but present on all immunoglobulins (with some exceptions which I need not discuss), are more primitive than both the class-specific determinants which correlate with the physical properties of the IgM and IgA molecules, and the allotypic systems which are confined to them. Thus the order of evolution would be duplication to form light and then quadruplication to form heavy chains, followed by the development of the IgG allotypic systems on them. Finally, differentiation of the heavy chain classes and development of the class-specific allotypic systems would occur.

I should like to close by drawing attention to one other type of reaction, conventionally included as immunological, which may well have great possibilities for taxonomy, and this is the mixed normal cell reaction. If under appropriate experimental conditions one mixes together mammalian lymphocytes of different genotypes, a mutual stimulation occurs and blast cells are formed. Cooper (1967) has shown mutual stimulation between the cells from two species of lamprey which are virtually indistinguishable in the ammocete or larval stage, even though

individuals do not readily stimulate one another. Immunologists are getting a surfeit of lampreys at the moment! This shy and unambitious creature has often to leave its native mud to bathe in the pure ichor of research. This phenomenon can be quantitated, and it seems possible that its application to the taxonomy, especially of lower organisms, might be of very great valne.

REFERENCES

Allison, A. C. and Blumberg, B. S. (1961). An isoprecipitation reaction distinguishing human serum-protein types. *Lancet* i, 634–7.

Cooper, A. J. (1967). Immunological responsiveness in ammocetes. Thesis presented for the degree of Ph.D. University of Birmingham.

Hill, R. L., Delaney, R., Fellows, R. E. Jr. and Lebovitz, H. C. (1966). The evolutionary origins of the immunoglobulins. *Proc. Natn. Acad. Sci., U.S.A.* **56**, 1762–9.

Kelus, A. S. and Gell, P. G. H. (1967). Immunoglobulin allotypes of experimental animals. *Prog. Allergy* **11**, 141–79.

Kirsch, J. A. W. and Poole, W. E. (1967). Serological evidence for speciation in the grey kangaroo, *Macropus giganteus* Shaw 1790 (Marsupialia: Macropodidae). *Nature, Lond.* **215**, 1097–8.

Manski, W., Halbert, S. P., Javier, P. and Auerbach-Pascal, T. (1967). On the use of antigenic relationships among species for the study of molecular evolution. *Int. Archs. Allergy appl. Immun.* **31**, 529–45.

Oudin, J. (1956). The specific precipitation reactions between blood of animals of the same species. *C.r. hebd. Séance. Acad. Sci. Paris* **242**, 2606–09.

Pollara, J., Finstad, J. and Good, R. A. (1966). The phylogenetic development of immunoglobulins, pp. 88–97. *In* Smith, R. T., Miescher, P. A. and Good, R. A. (eds.) "The Phylogeny of Immunity". University of Florida Press, Gainesville, Pla.

Stanworth, D. R. and Henney, C. S. (1964). "Protides of the Biological Fluids" p. 155. Elsevier, Amsterdam.

8. Serological Assessment of Relationships in a Flowering Plant Family (Solanaceae)

J. G. HAWKES and W. G. TUCKER

Department of Botany, University of Birmingham, England

ABSTRACT

The first stage of a serotaxonomic revision of the family Solanaceae is described in which seeds have been used as the source of saline soluble protein. Two contrasting classifications of the family are outlined and a preliminary survey of the serological relationships between 15 genera is briefly referred to. The scoring of morphological characters is also mentioned.

Examples are given of the way in which the experimental method is providing valuable data on serological relationships within *Datura* and within *Solanum*; these are followed by a discussion of the systematic position of certain genera closely similar to *Solanum*. The serological affinity of *Cestrum* is then discussed. Finally, the little understood boundaries of Solanaceae with Scrophulariaceae and other families are dealt with. The controversial position of the solanaceous genus *Schizanthus* is outlined and illustrated by experimental results. Serological data show a somewhat distant relationship of *Schizanthus* to the rest of the genera investigated.

The taxonomic significance of serological results is discussed.

INTRODUCTION

Plant serology has successfully been employed as a taxonomic aid in this department in the study of tuber-bearing solanums (Gell *et al.*, 1960). Three years ago facilities were provided for an extension of the work to cover the whole Solanaceae, utilizing seed rather than tuber proteins, and a preliminary survey was started to see if the method was suited to an investigation on a family scale.

The Solanaceae, a family of some 98 genera, is badly in need of taxonomic revision, as we shall see shortly. It is a family of great commercial importance, containing many food- and drug-producing plants. Figure 1 shows the arrangement of the genera according to Bentham and Hooker (1876) and according to Wettstein (1895) in Engler and

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 77–98.

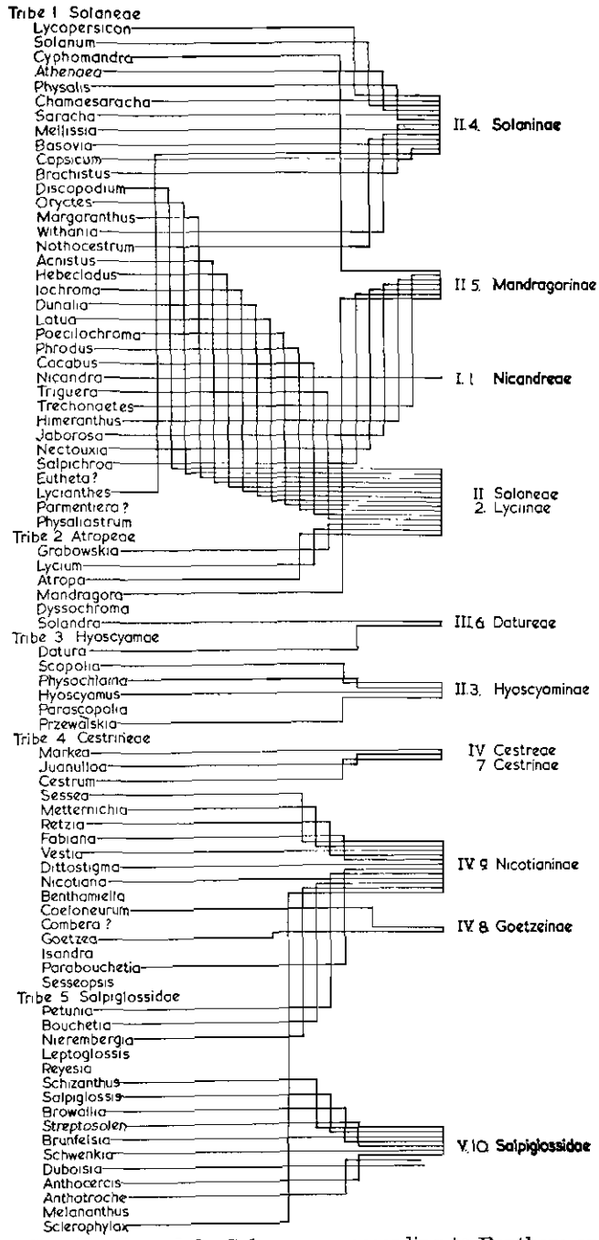


FIG. 1. The Classification of the Solanaceae; according to Bentham and Hooker (1876) on the left, and according to Wettstein (in Engler and Prantl, 1895) on the right.

Prantl. These two systems are the most comprehensive and detailed classifications of the family. Although both systems utilize much the same floral characters to delineate the tribes, and even allowing for the fact that one system arranges the genera in five tribes and the other system in ten tribes, there are a great number of differences between the two. Part of the present work is to assess the relative value of the two classifications and to use the serological data, together with other information, in establishing a more satisfactory and natural classification for the family.

Seed harvested from plants grown at the Botanic Garden was used as the source of saline-soluble antigenic protein. The seeds contained oils and alkaloids which were removed during the preparation of the extract. A certain amount of seed was ground in a mill and then used to prepare an oil-free and water-free "acetone powder" (Nason, 1955). Saline-soluble protein was extracted by shaking the powder with 0·54 Molar phosphate buffered saline at pH 7·5 for 30 min and the supernatant was removed by centrifuging. To purify, it was removed from solution by adding ammonium sulphate to saturation. The protein was redissolved in saline, dialysed against saline overnight and then centrifuged to clarify. The extracts were then free of oil and alkaloid. The protein content was estimated colorimetrically by the production of biuret (Gornall, 1949) and adjusted to 0·2%, by dilution or concentration (the latter by dialysis against Carbowax, see Kohn, 1959). Sodium azide was used as a bacteriocide and those extracts not required immediately were frozen.

Pairs of antisera were raised in New Zealand white rabbits. Initially the injections were intramuscular and incorporated Freund's complete adjuvant (Freund and Bonanto, 1944). Later on, 0·2% protein solution was given intravenously.

The main bleed was taken when the breathing rate following injections approached 380 per min. The duplicated antisera matched each other exactly in content, and were very close in strength. The antisera which were raised to different genera were also all of similar strength.

Antigen–antibody interaction was studied by double diffusion in agar gel, using 0·8% Ionagar No. 2 veronal buffered at pH 8·6. The double diffusion plates were dried and stained with Ponceau S. We should like to record our grateful thanks to Miss S. Marsh for growing the plants and Mrs. A. Ferguson for technical assistance during the course of this work.

PRELIMINARY SURVEY

A preliminary serological survey of 15 genera was first carried out, using double-diffusion techniques in agar gel. The terms used by

Ouchterlony (1960) have been followed in the details of analysis of the plates. The results indicate a number of strong serological relationships and certain weak relationships between the genera investigated. By the phrase "serological relationship" we mean the ability of proteins or other similar substances of one species to form a series of precipitin lines in the agar against the antiserum raised to those of another species. The degree of relationship is judged by the intensity of the lines and their number. We are, therefore, in fact judging the ability of various

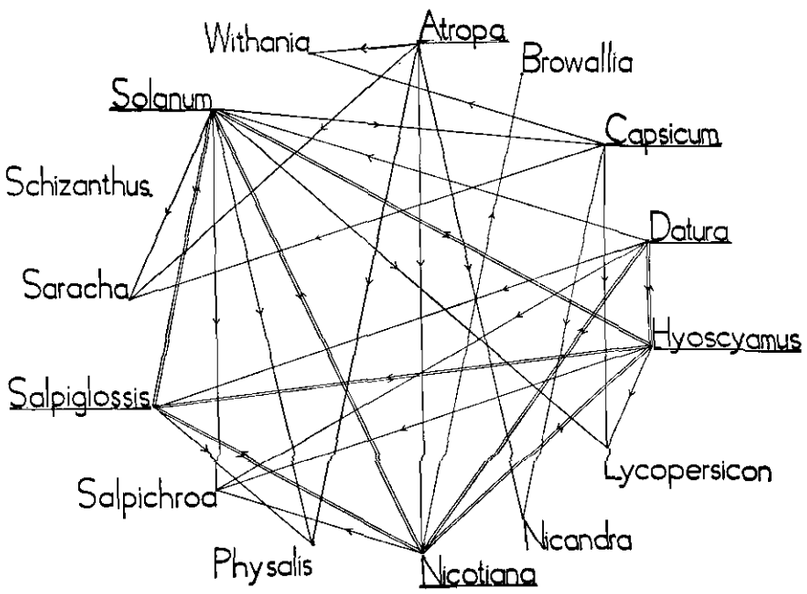


FIG. 2. Results of the preliminary serological survey. Antisera were raised to those genera underlined. For further explanation see text.

species to produce identical or closely similar proteins or other antigenic substances, and the tentative inference (to be verified by reference to other information) is that this serological relationship is indicative of a genetic or evolutionary relationship. This, of course, will be a matter for discussion at the present conference.

The preliminary survey is too lengthy to describe in detail. A summary of some of the results is shown in Fig. 2.

Species from 15 genera were extracted, and antisera were raised to 7 of them (underlined in the figure). Cross-reactions between each antiserum and the other 14 antigenic extracts were studied, and arranged

in a list of order of closest similarity to the homologous reaction. The cross-reactions which were closest to the homologous reaction are shown in the diagram in the form of an arrow pointing to the antigenic extract concerned. The double lines present between genera to which antisera were raised indicate that strong cross-reactions occurred in both directions. The diagram shows strong serological relationships between the genera *Datura*, *Hyoscyamus*, *Nicotiana*, *Salpiglossis* and *Solanum* since they are connected to each other by double lines. In fact they possess a total of from seven to twelve "connecting lines" to other genera. We would interpret this as indicating that, of the genera tested, these form a kind of "central core" in the family, related strongly to each other, and in various ways related to other "peripheral" genera. At the other end of the scale lie *Capsicum* and *Atropa*, with five connections only. Apart from the group to which antisera were raised and which stand a better chance of connecting, both as antisera and as antigens, we have the other group of genera which were used as antigens only, not as antisera. Of these, *Salpichroa* shows 4 connections, *Physalis*, *Saracha* and *Lycopersicon* 3, *Nicandra* 2, *Browallia* 1 and *Schizanthus* 0.

Clearly, more verificatory work needs to be done in this sphere but we have shown enough, we hope, to give some indication of the lines along which we are working in this broad general survey.

We shall now continue with some examples of how the experimental methods are providing data on species relationships within a genus and the arrangement of genera into tribes, as well as the probable boundaries of the family with its neighbours.

EXAMPLES OF RESULTS

(a) Species relationships within a genus

Our first example shows the type of cross-reaction within a genus that we have come to accept as normal. It concerns two species and two varieties of *Datura*, and an antiserum to one of the species, *D. stramonium*. This latter has an erect spiny fruit and the flowers are white and trumpet-shaped. Variety *tatula* is the same, except that the flowers are purple. Variety *inermis* has a purple flower, but the fruit lacks spines. *Datura inoxia* shows a number of differences: the flowers are much longer, the plant is very pubescent, and the fruit is pendulous. Extracts were made of the seed of these four plants, and an antiserum was raised to *D. stramonium*. The homologous reaction, indicated by the arrows in Fig. 3(*a*), is seen to consist of four lines. The two varieties of *D. stramonium*, *inermis* and *tatula*, produce a cross-reaction differing very little from the homologous one. The extract of *D. inoxia*, however, differs significantly in that line 2 is missing.

This clearly demonstrates that in *Datura*, small variations at varietal level are not accompanied by any great change in the saline-soluble seed proteins, while large differences involving several characters between two taxa classed as different species are apparently accompanied by a clear change in the seed protein content.

(b) Range of cross-reactions within *Solanum*

The genus *Solanum* is the largest in the family, with at least 2 000 species. It shows a wide morphological range in habit, leaf form, pubescence and armament, though the corolla and fruit form remain relatively constant.

Some of the types of *Solanum* growing at the Birmingham University Botanic Garden are as follows (classification based on Seithe, 1962):

A. Subgenus *Solanum*
   1. Sect. Solanum (=Morella). *S. nigrum*.
      This has small stellate flowers and soft pulpy berries.
   2. Sect. Pseudocapsicum. *S. hendersonii*.
      The species of this section possess brightly coloured berries similar to those of *Capsicum* but in other respects are clearly identifiable as solanums.
   3. Sect. Jasminosolanum. *S. seaforthianum*.
      This is rather similar to *S. dulcamara*, with stellate purple flowers, divided leaves and vine-like habit.
   4. Sect. Tuberarium. *S. phureja*.
      This is a cultivated tuber-bearing species with pinnate leaves and an articulated pedicel.

B. Subgenus *Archaesolanum*
   S. simile. This is an example of the Australian group of species with a rotate corolla and leaves with strap-like segments or sometimes with an entire lamina.

C. Subgenus *Stellatipilum*
   1. Sect. Brevantherum. *S. mauritianum*.
      This species is characterized by a covering of woolly stellate hairs but no spines.
   2. Sect. Androceras. *S. heterodoxum*.
      The stellate hairs are not so pronounced and needle-like spines are developed. There is a slight tendency to zygomorphy in the corolla.
   3. Sect. Oliganthes. *S. topiro*.
      This species also possesses stellate hairs, has a woody stem, stellate white flowers and firm red berries.

Extracts have recently been made of all these types, and some are

illustrated in Fig. 3(*b*), where the homologous reactions are again indicated by arrows. The cross-reactions are very varied but consist of at least two spurs in all cases, so the much wider morphological range is accompanied by great variation in the seed proteins.

Our preliminary results thus indicate that some of the differences between sections in the genus *Solanum* are serologically equivalent to inter-generic differences elsewhere in Solanaceae. This lends strength to the argument that *Solanum* ought to be divided into several smaller genera on the basis of leaf and pubescence characters.

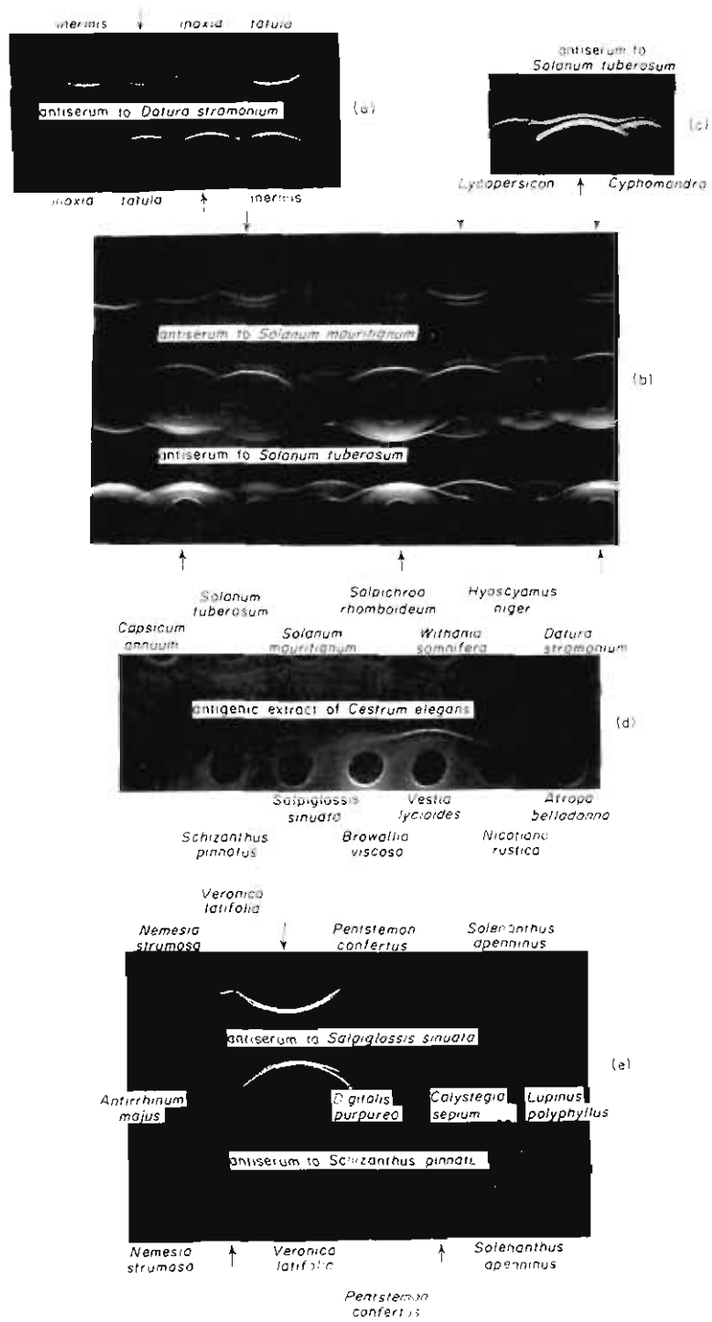
(*c*) The inter-relationship of certain genera and their systematic position in Bentham and Hooker's and in Wettstein's classification

(i) In Fig. 1 it is seen that Bentham and Hooker place *Lycopersicon*, *Solanum* and *Cyphomandra* together at the beginning of the list in the tribe Solaneae. In Wettstein's system, however, *Cyphomandra* is placed in another group. This is because the stamen filament is attached to the back of the anthers, instead of to the base. Though the corolla structure is very similar to that of *Solanum*, *Lycopersicon* is separated from *Solanum* by virtue of the anther dehiscence and the sterile anther tip, but has often been included in it. When extracts of these plants are plated out next to the homologous reaction of *Solanum tuberosum*, as shown in Fig. 3(c), the cross-reactions are found to be very strong. *Lycopersicon* produces some lines of identity, while *Cyphomandra* produces a number of very strong spurs. The three genera are thus shown to possess closely similar saline-soluble seed proteins.

This close serological relationship of the genera *Solanum*, *Lycopersicon* and *Cyphomandra* agrees best with Bentham and Hooker's system.

(ii) The genus *Cestrum* has been separated from *Nicotiana* and *Vestia* by Wettstein because the fruit is a berry, while the other two genera possess capsules. In *Cestrum*, the style and stamens are not exserted, whilst in *Nicotiana*, the filament of one stamen is frequently shorter than the rest, and in *Vestia* the style and stamens are strongly exserted. Bentham and Hooker have placed all three in the same tribe because of their otherwise overall similarity. In the experimental work, antisera have been raised to species from 13 different genera from all parts of the family, including *Nicotiana* and *Vestia*.

In Fig. 3(*d*), the trough contains an antigenic extract of *Cestrum elegans*, while the cups each contain a different antiserum. The cross-reactions with *Cestrum* are faint except with the antiserum to *Vestia*, indicating that the proteins of these two taxa have an overall similarity. If this is taken to indicate a close relationship between the two, then again it is contrary to Wettstein's scheme.


inermis
inoxia
tatula
antiserum to Datura stramonium
(a)
antiserum to Solanum tuberosum
(c)
Lycopersicon
Cyphomandra
inoxia
tatula
inermis
antiserum to Solanum mauritianum
(b)
antiserum to Solanum tuberosum
Solanum tuberosum
Salpichroa rhomboideum
Hyoscyamus niger
Capsicum annuum
Solanum mauritianum
Withania somnifera
Datura stramonium
antigenic extract of Cestrum elegans
(d)
Salpiglossis sinuata
Vestia lycioides
Atropa belladonna
Schizanthus pinnatus
Browallia viscosa
Nicotiana rustica
Veronica latifolia
Nemesia strumosa
Penstemon confertus
Solenanthus apenninus
antiserum to Salpiglossis sinuata
(e)
Antirrhinum majus
Digitalis purpurea
Calystegia sepium
Lupinus polyphyllus
antiserum to Schizanthus pinnatus
Nemesia strumosa
Veronica latifolia
Solenanthus apenninus
Penstemon confertus

(d) Serological relationships between Solanaceae and other flowering plant families

The Scrophulariaceae is generally considered by taxonomists to be the family most nearly related to the Solanaceae. Although the flowers of Scrophulariaceae are mostly zygomorphic, whilst those of Solanaceae are actinomorphic for the most part, there are certain exceptions. Thus, *Schizanthus* is a solanaceous genus with highly zygomorphic flowers, whilst those of *Verbascum* in Scrophulariaceae are almost actinomorphic. *Schizanthus* is of particular interest to us, not only because of its zygomorphic flower, but because in the general survey it showed no strong serological relationships to any other solanaceous genus tested. It had been at first classified in Scrophulariaceae, but when it was discovered that the "typical" solanaceous genera possessed internal phloem and that the wall dividing the carpels was set obliquely to the floral axis, these two criteria were used to separate Solanaceae from Scrophulariaceae. Because *Schizanthus* also possessed these two features it was re-classified into Solanaceae.

As we have seen, of all the extracts made of the seed of Solanaceae, *Schizanthus* alone fails to produce a good cross-reaction with any solanaceous antiserum.

We therefore began to wonder whether it would show more serological relationships to Scrophulariaceae and whether this latter and even other families might show cross-reactions to solanaceous antisera. It became important to know what, in fact, were the serological boundaries of Solanaceae.

Antisera were raised to *Schizanthus* and *Salpiglossis*, and extracts

---

FIG. 3. Double-diffusion plates showing cross-reactions with antisera raised to selected species in the Solanaceae. The homologous reactions are indicated by arrows. Antisera are normally placed in the troughs and antigens in a series of cups on either side.

(*a*) Cross-reactions within *Datura*.

(*b*) Cross-reactions within *Solanum*.
Cross-reactants as follows. Top row, left to right: *S. hendersonii*, *S. tucumanense*, two unidentified species, *S. topiro*. All five species are representatives of the woody species, described in the text. The extracts in the bottom row are the same as the top row. The middle row contains examples of the Morella section and others. Left to right: *S. aethiopicum*, *S. tuberosum*, *S. mauritianum*, *S. nigrum*, *S. tuberosum*, *S. mauritianum*, *S. nigrum* var. *guineense*, *S. tuberosum*.

(*c*) Cross-reactions of *Lycopersicon* and *Cyphomandra* with *Solanum tuberosum*.

(*d*) *Cestrum elegans* tested against all the antisera.

(*e*) Cross-reactions with *Salpiglossis* and *Schizanthus*.

from species belonging to several different families were plated out against these and six other solanaceous antisera.

We chose Scrophulariaceae for obvious reasons, and also an example from Boraginaceae and Convolvulaceae, which are placed with Solanaceae and Scrophulariaceae in the order Tubiflorae. Finally, we chose *Lupinus*, a genus in Leguminosae, very far removed in all conventional classifications from Solanaceae. The extracts tested were as follows:

Antirrhinum majus	Scrophulariaceae
Digitalis purpurea	
Pentstemon confertus	
Nemesia strumosa	
Veronica latifolia	
Solenanthus apenninus	Boraginaceae
Calystegia sepium	Convolvulaceae
Lupinus polyphyllus	Leguminosae

All these extracts produced faint, extremely faint or no cross-reactions with the antisera. As examples, the reactions with the *Salpiglossis* and *Schizanthus* antisera are shown in Fig. 3(*e*).

The *Schizanthus* homologous reaction is much fainter than those of the other twelve antisera; however, three precipitin lines can be distinguished. The cross-reaction of *Schizanthus* with the *Salpiglossis* antiserum, and of *Salpiglossis* with the *Schizanthus* antiserum are both faint arcs, indicating that their saline soluble seed proteins are mostly different.

Lupinus extracts produced no cross-reactions and thus agreed with the distant relationship between Leguminosae and Solanaceae postulated by taxonomists. On the other hand, Solanaceae, Scrophulariaceae, Boraginaceae and Convolvulaceae are generally classed in the order Tubiflorae and should be expected to show some degree of relationship. In fact, *Calystegia* (Convolvulaceae) showed no cross-reaction with *Schizanthus*, whilst *Solenanthus* (Boraginaceae) showed only an extremely faint one with *Schizanthus*. Neither gave any cross-reaction with any of the other antisera.

The remainder of the cross-reactions are with extracts of the Scrophulariaceae and they are all faint. *Veronica* shows no cross-reaction with the *Salpiglossis* antiserum and *Nemesia* produces no reaction with the *Schizanthus* antiserum.

The results indicate that there are at least three types of cross-reacting extracts present here, i.e. *Schizanthus*, *Salpiglossis* and Scrophulariaceous genera. None of them cross-reacts strongly with the others, but the fact that they produce faint cross-reactions shows that they have something in common, and must be to some extent serologically related.

DISCUSSION

The results we have obtained so far are, we think, very promising. We have used simple double-diffusion techniques in the present work because these methods enable us to sketch in a broad general picture. In more detailed studies at the species level or below, the more delicate techniques of parallel immuno-electrophoresis are clearly more relevant, as the work of Lester (1965) and Hawkes and Lester (1966) has shown for the tuber-bearing species. We plan to use such methods also in the present work when the main outlines have been completed and we can turn to detailed studies in certain selected genera, notably in other areas of the genus *Solanum*.

By and large, our serological information correlates well with morphological characters and sometimes lends weight to Bentham aud Hooker's classification and sometimes to that of Wettstein.

In all newly discovered or applied techniques in taxonomy there is a strong temptation to extol them in the first flush of enthusiasm as of greater importance and thus able to provide characters of greater weight than ones hitherto in use. This temptation must be, of course, firmly resisted. Even though it might be said that we are dealing with proteins which could be considered as a direct expression of the genotype, serological data are nevertheless characters worthy of taxonomic consideration but of not necessarily greater importance than other characters, as Fairbrothers (1966) has pointed out.

Serological information should therefore be used as one more independent criterion to help provide a worthwhile classification based on maximum correlation of characters and should not be given *a priori* weighting at the expense of other characters. Having sounded a note of caution, we are nevertheless convinced that the perspectives for the future in serological work are vast and immensely promising. At one end of the scale one can establish fine differences between species or infraspecific taxa, and at the other one can even begin to elucidate relationships between families.

REFERENCES

Bentham, G. and Hooker, J. D. (1876). "Genera Plantarum." Reeve, London.

Fairbrothers, D. E. (1966). The comparison and interpretation of serological data in plant systematics. *In* Landa, Z. (ed.) Mechanism of mutation and inducing factors. "Proc. Sym. Mutation Process", 458–64. Czechoslovak Acad. Sci., Prague.

Freund, J. and Bonanto, M. V. (1944). *J. Immun.* **48**, 325–34.

Gell, P. G. H., Hawkes, J. G. and Wright, S. T. C. (1960). The application of immunological methods to the taxonomy of species within the genus *Solanum*. *Proc. Roy. Soc. B.* **151**, 364–83.

Gornall, A. G., Bardawill, C. J. and David, M. M. (1949). Determination of serum proteins by means of the biuret reaction. *J. biol. Chem.* **177**, 751–66.

Hawkes, J. G. and Lester, R. N. (1966). Immunological studies on the tuber-bearing Solanums. II. Relationships of the North American species. *Ann. Bot.* N.S. **30**, 269–90.

Kohn, J. (1959). A simple method for the concentration of fluids containing protein. *Nature, Lond.* **183**, 1055.

Lester, R. N. (1965). Immunological studies on the tuber-bearing Solanums. I. Techniques and South American species. *Ann. Bot.* N.S. **29**, 609–24.

Nason, A. (1955). *In* Colowick, S. P. and Kaplan, N. O. (eds.) "Methods in Enzymology", Vol. 1, 62–63. Academic Press, New York.

Ouchterlony, O. (1960). Interpretation of comparative immune precipitation patterns obtained by diffusion in gel techniques. *In* Heidelberger and Plexia (eds.) "Immunochemical Approaches to Problems in Microbiology", 5–19. Rutgers University Press, New Jersey.

Seithe, A. (1962). Die Haararten der Gattung Solanum L. und ihre taxonomische Verwertung. *Bot. Jb.* **81**, 261–336.

Wettstein, R. (1895). *In* Engler and Prantl (eds.) "Die natürlichen Pflanzenfamilien", **4**, 3, 4–38. Engelmann, Leipzig.

9. Serotaxonomy in Ranunculaceae

UWE JENSEN

Institut für Pharmakognosie, University of Kiel, West Germany

(Extended summary)

For a long time the main classification of the family Ranunculaceae was based essentially on fruit characters. The primary distinction was between genera with multiovular follicles and those with uniovular achenes, apart from *Paeonia*, which is now generally removed from Ranunculaceae altogether.

The first karyological works by Langlet (1932) and Gregory (1941) were the beginning of an extensive new examination of the systematics of this family.

We have undertaken the examination of serological characters and for this purpose we have restricted our work to the proteins of ripe seeds at the time of dormancy. The method used was the presaturation (or absorption) technique, which is well known from the literature. Such a technique allows comparison of two antigenic systems (for example, X and Y) with respect to an antiserum (for example against A, *Caltha palustris*), and shows which of these antigenic systems have the greatest similarity to the antiserum (and hence the antigenic system of *Caltha palustris*). In subsequent experiments we compared our antigenic systems with anti-B, anti-C, anti-D etc., instead of anti-A (i.e. anti-*Delphinium*, anti-*Adonis*, anti-*Clematis*, etc. instead of anti-*Caltha*).

The important question is how to interpret the serological data. From all these individual results we can build up a system of the serological similarities between the genera of the Ranunculaceae (Fig. 1) (see also Jensen, 1966). I wish to emphasize, however, that this is not a phylogenetic system based on a single character or a group of characters. In the first place, we have no evidence as to what is primitive and what is derived. This means, in effect, that we do not know the starting point of our system. If we based our system on non-serological

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 89–91.

characters, we should place the starting point between *Trollius*, *Cimicifuga* and *Caltha*, and should arrive at the scheme shown in Fig. 2 (Jensen, 1966).

It is interesting to note that on the whole the agreement between systems based on serological and non-serological data is rather high. In spite of this, the system shown in Fig. 2 cannot yet be the natural system of relationship, because we have only considered some other

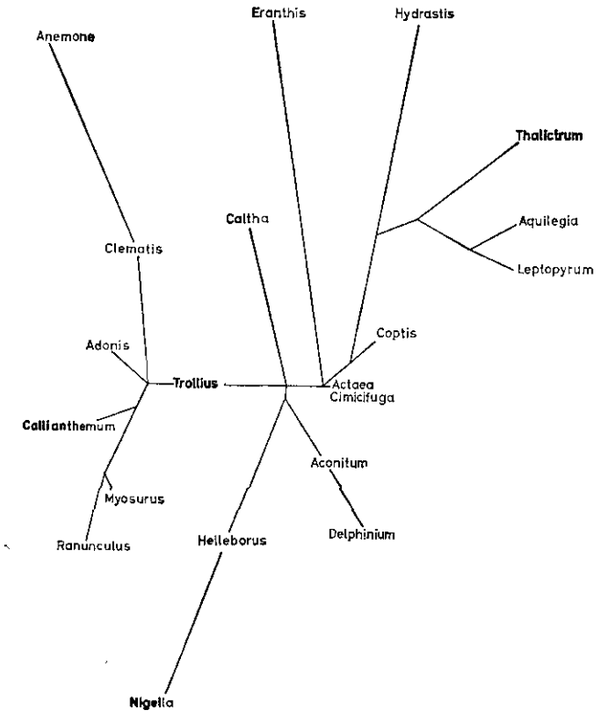


FIG. 1

aspects of the "primitive" genera.

To provide a real contribution to systematics, serotaxonomie work should be regarded as a biological task which should draw in and evaluate all available characters, without over-emphasizing the serological data at the expense of the others. It should thus examine the relationship of plants and animals in their entirety.

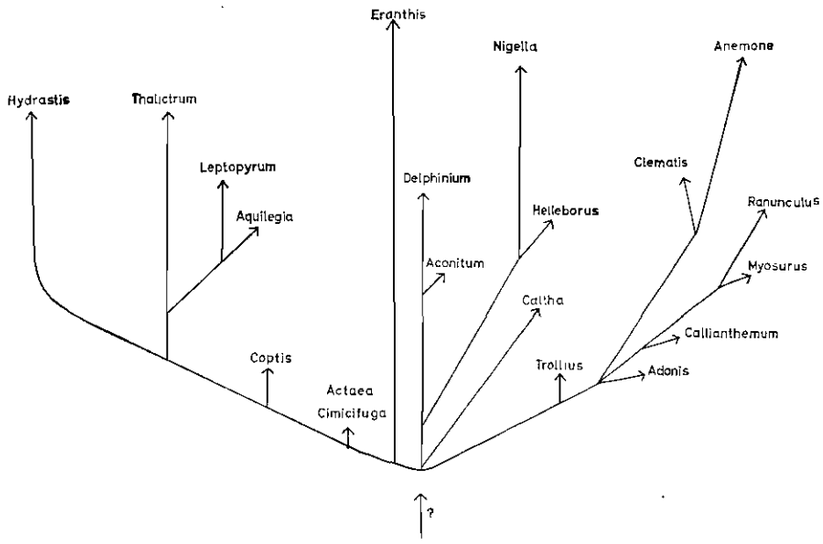


FIG. 2

REFERENCES

Gregory, W. C. (1941). Phylogenetic and cytological studies in the Ranunculaceae Juss. *Trans. Am. phil. Soc.*, N.S. **31**, 443–521.

Jensen, U. (1966). Die Verwandtschaftsverhältnisse innerhalb der Ranunculaceae aus serologischer Sicht. *Ber. dt. bot. Ges.* **79**, 407–12.

Langlet, O. (1932). Uber Chromosomenverhältnisse und Systematik der Ranunculaceae. *Svensk bot. Tidskr.* **26**, 381–440.

10. Variability of Proteins I and II in the Seeds of Species of the Genus *Phaseolus*

JOSEF KLOZ and EVA KLOZOVA

Institute of Experimental Botany, Czechoslovak Academy of Sciences, Prague, Czechoslovakia

ABSTRACT

Distribution of proteins I and II was studied in *Phaseolus*. These proteins were found in species related to *P. vulgaris*. Transitional types were found in cultivars of *P. coccineus*, whilst in *P. vulgaris* polymorphism between and within cultivars as well as unusual mutations in protein specificity were established. The behaviour of both proteins in interspecific and intraspecific hybrids was also studied.

In this paper we shall discuss briefly the variability of two individual seed proteins of the genus *Phaseolus* which we have studied by serological methods and especially by immuno-electrophoresis.

These two proteins belong to the seed albumin fraction of *P. vulgaris* and were obtained from seed extracts (cotyledons) of *P. vulgaris* cv. Veltruská Saxa by ammonium sulphate fractionation at between 30–50% saturation. We designated the protein moving to the anode at pH 7·6 as *Phaseolus* protein I and the protein moving to the cathode as *Phaseolus* protein II (see Fig. 1).

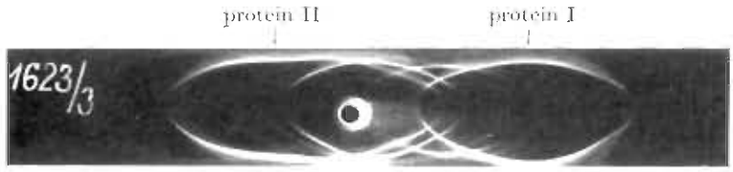


FIG. 1. Protein I (on anode side) and protein II (on cathode side) in the extracts of seeds of *P. vulgaris* cv. Veltruská Saxa detected by antiserum against the seed albumin fraction of *P. vulgaris* cv. Veltruská Saxa. The antiserum was prepared by immunization of rabbits by means of proteins prepared by fractionation with ammonium sulphate at between 30–50% saturation. This antiserum was used in all other cases.

Systematics Association Special Volume No. 2, "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 93–102.

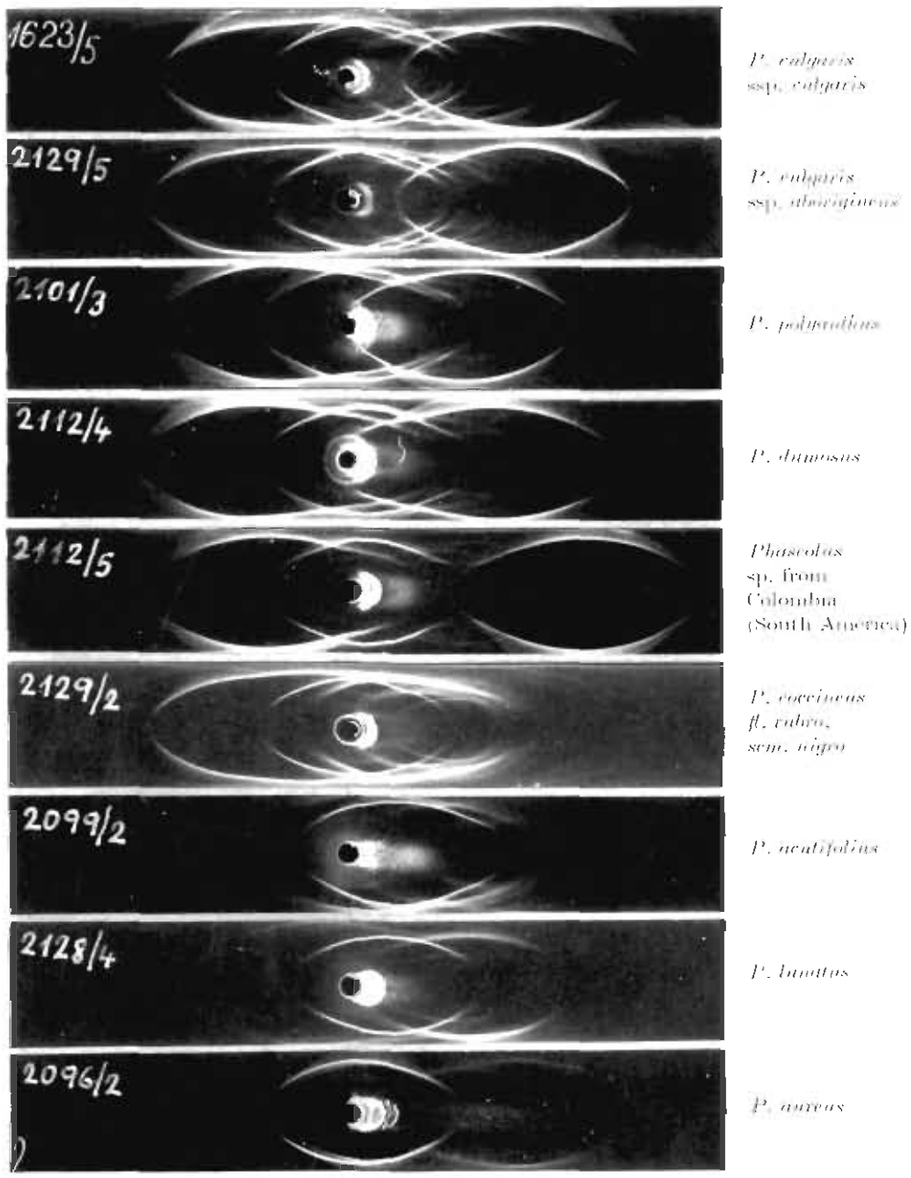


FIG. 2. Immuno-electrophoretograms showing the occurrence of proteins I and II in several different species of the genus *Phaseolus* related to *P. vulgaris*.

We shall now consider the distribution of both proteins in different species of the genus *Phaseolus* and in different cultivars of *P. vulgaris* and *P. coccineus* as well as their inheritance in hybrids between these species.

Firstly we shall consider protein I, which is expressed strongly in the seeds of *P. vulgaris* ssp. *vulgaris*, *P. vulgaris* ssp. *aborigineus*, *P. polyanthus*, *P. dumosus* and *Phaseolus* sp. (undescribed species from Bogotá, Colombia), but is expressed less strongly in *P. coccineus*. All these are species which we suppose to be inter-related. Protein I is absent from other species, such as *P. acutifolius*, *P. lunatus*, *P. polystachyus*, *P. atropurpureus*, *P. aureus*, etc. (see Fig. 2).

Protein I is also lacking in other investigated representatives of the tribes Phaseoleae, Viciceae and Hedysareae. Therefore it could be said that the protein I is specific for the group of species related to *P. vulgaris*.

During subsequent work we confirmed our former findings that protein I is present in varying degrees in different cultivars of *P. coccineus*, whilst in some it is completely absent. Figure 3 shows some

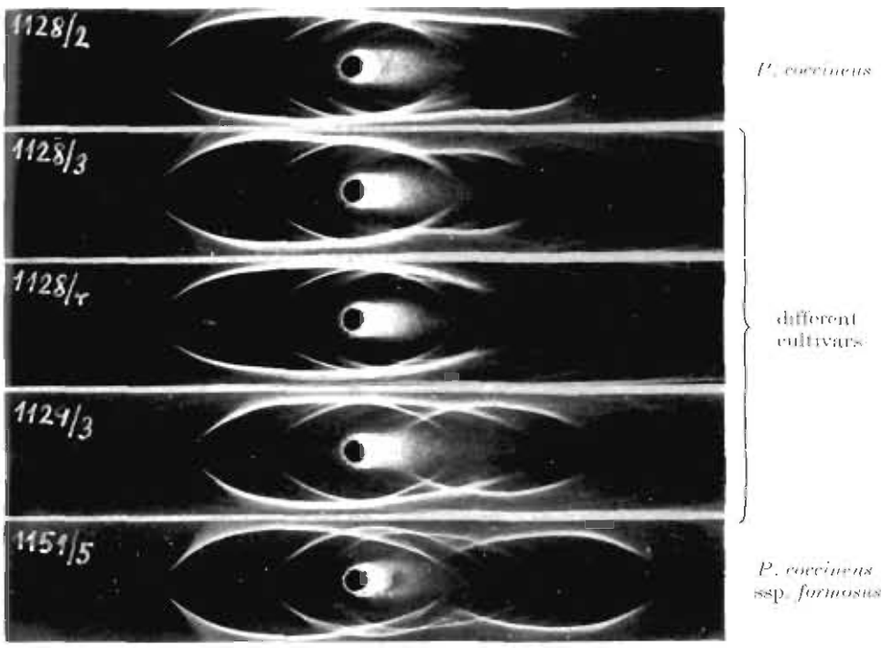
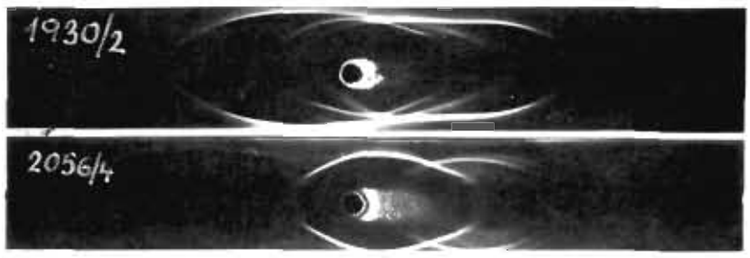


Fig. 3. Several examples of cultivars of *P. coccineus* (selected from 22 cases studied) with differently developed protein I (latent transitions).

examples of this variation, based on immuno-electrophoretic analysis of single seeds of different cultivars. A whole range in protein characters can be seen between the species *P. vulgaris* and *P. coccineus*. We suppose that these transitional types are either surviving intermediate stages between two taxa which have diverged from a common ancestor or are due to recent or ancient hybridization.



cv. Los Angeles I

cv. Krupnaja Sacharnaja

FIG. 4. Two exceptions, where protein I is modified or absent, found in 659 *P. vulgaris* cultivars: cv. Los Angeles I – protein I seems to indicate hybrid origin; cv. Krupnaja sacharnaja – protein I is wholly lacking.

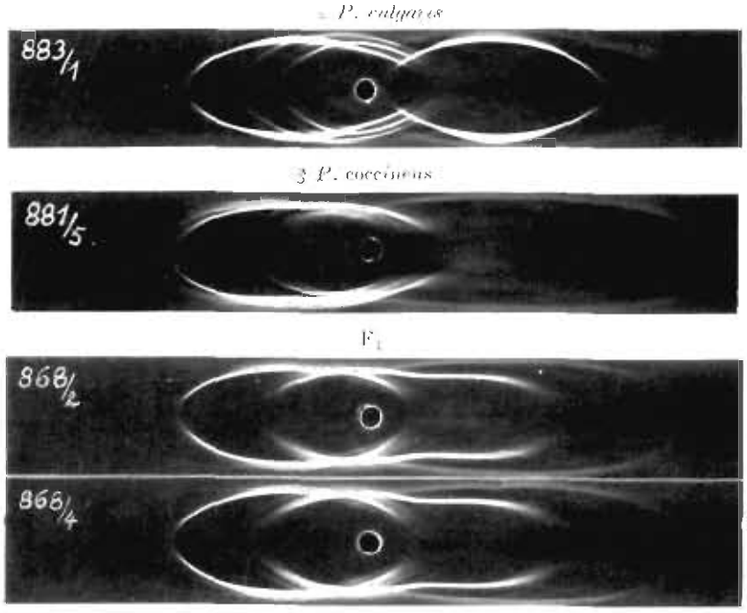


FIG. 5. Inheritance of protein I in hybrids of *P. vulgaris* cv. Veltruská Saxa × *P. coccineus flore rubro semine nigro*.

F_2

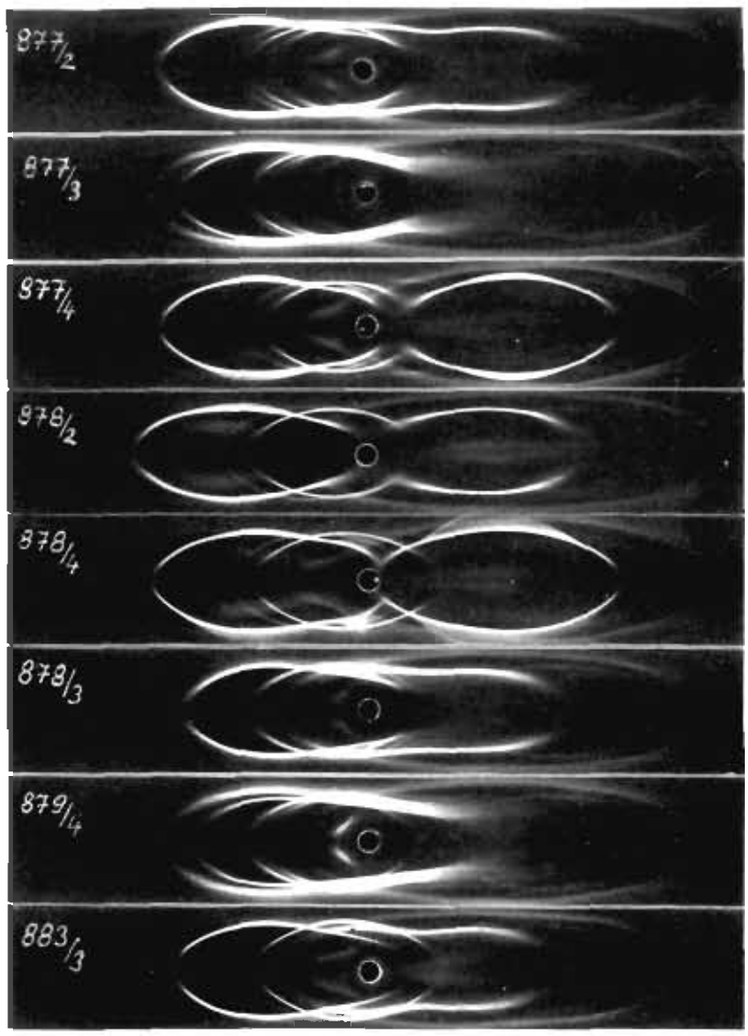


FIG. 5 (contd.). For legend see facing page.

We have also analysed 659 cultivars of the species *P. vulgaris*. Protein I is very evident in all but two of the cultivars examined (see Fig. 4). In one cultivar it is completely absent and we cannot explain this otherwise than as a loss mutation of a protein character. In the other cultivar protein I was not fully developed: when we analysed this

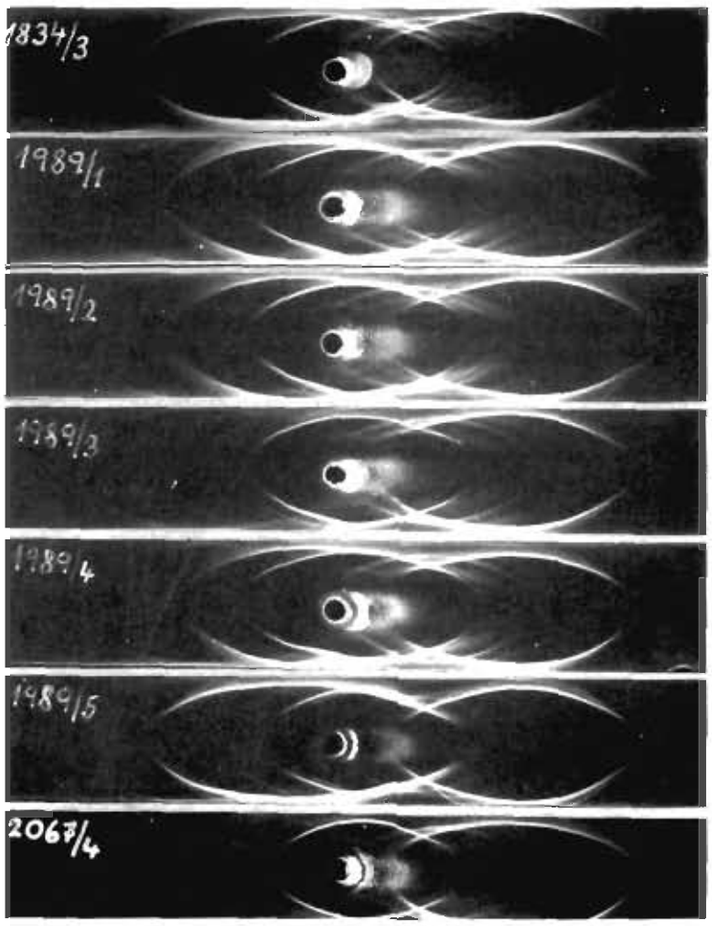


FIG. 6. Several examples showing polymorphism in Protein II found in a survey of 659 cultivars of *P. vulgaris*

case in more detail we found again that different transitional forms of this character were present in single seeds of the cultivar. We suppose that in this last case the variability could be due to hybridity.

It can be concluded that protein I is specific for *P. vulgaris* and closely related species but that it is also possible to find an interesting exception to this amongst 659 lines examined.

We also analysed protein I in hybrids between *P. vulgaris* and *P. coccineus*. This protein is of intermediate type in the F_1 generation

and segregates in the F_2 not only to the pure parental types but also to many transitional types between both parents (see Fig. 5). This protein character may thus be polygenically controlled.

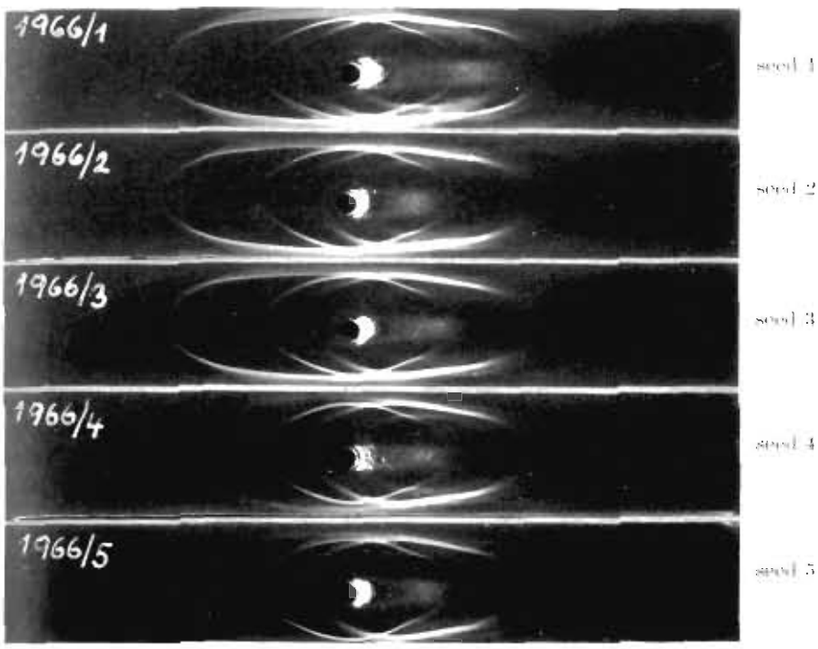


FIG. 7. Example of polymorphism of protein II in *P. coccineus* var. *albonanus* cv. Désirée. This is an example of the variation between single seeds of the cultivar from a population.

We shall next consider protein II (see Fig. 1).

When we studied the distribution of this character in our species collection of *Phaseolus*, we again found that it occurred only in the species related to *P. vulgaris* (i.e. *P. vulgaris* ssp. *vulgaris*, *P. vulgaris* ssp. *aborigineus*, *P. polyanthus*, *P. dumosus*, *Phaseolus* sp. from Bogota and *P. coccineus* (see Fig. 2).

Further analyses of single seeds of *P. vulgaris* cultivars have shown that in most cases protein II is present but in many it is wholly lacking. To be exact, out of 659 cultivars studied it is present in 611, but lacking in 48, i.e. 7·26%. We have not yet found any correlation between this protein character and other characters, and we suppose that this must be a simple case of polymorphism (see Fig. 6).

The Czechoslovak cultivar Dětenická of the species *P. vulgaris* was obtained by means of individual selection from the cultivar Präsident Roosevelt (former name for the cultivar Kapitän Weddingen). However, according to our analyses protein H is absent from the cultivar Präsident Roosevelt but present in the cultivar Dětenická. For this reason we were obliged to suppose that cv. Präsident Roosevelt was not homogeneous in regard to protein H. We analysed single seeds from samples of different origin but could not find any polymorphism in protein H in this cultivar. Finally, we found the expected polymorphism in a sample of seeds sent by the firm Vatter of Bern. We analysed 98 single seeds of this sample altogether: 71 seeds possessed protein H 26 seeds lacked it and in one case the protein was of intermediate form. The ratio was 2·8:1. As we have said before, we consider this to be a

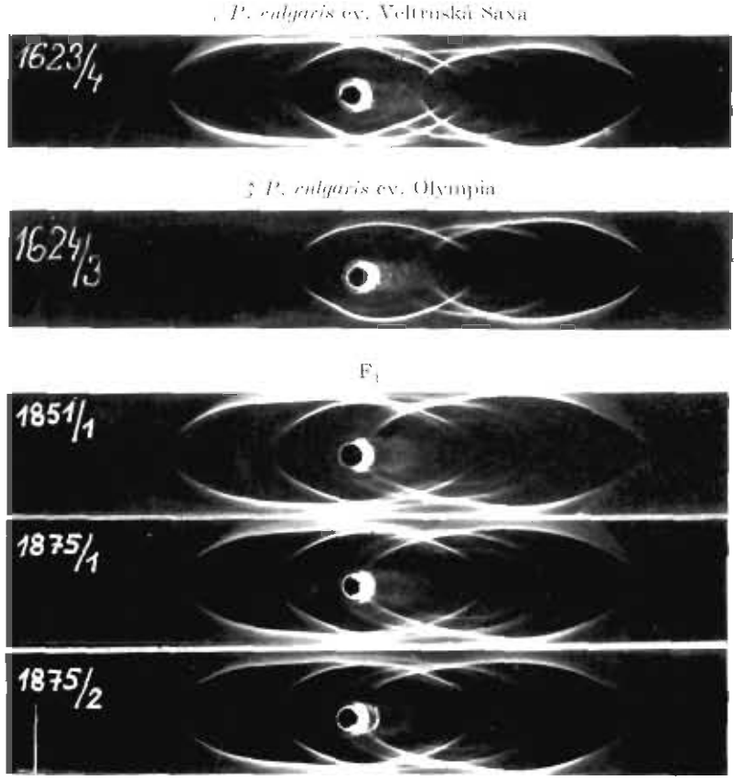


Fig. 8. Inheritance of protein H in hybrids of P. vulgaris cv. Veltruská Saxa × *P. vulgaris* cv. Olympia.

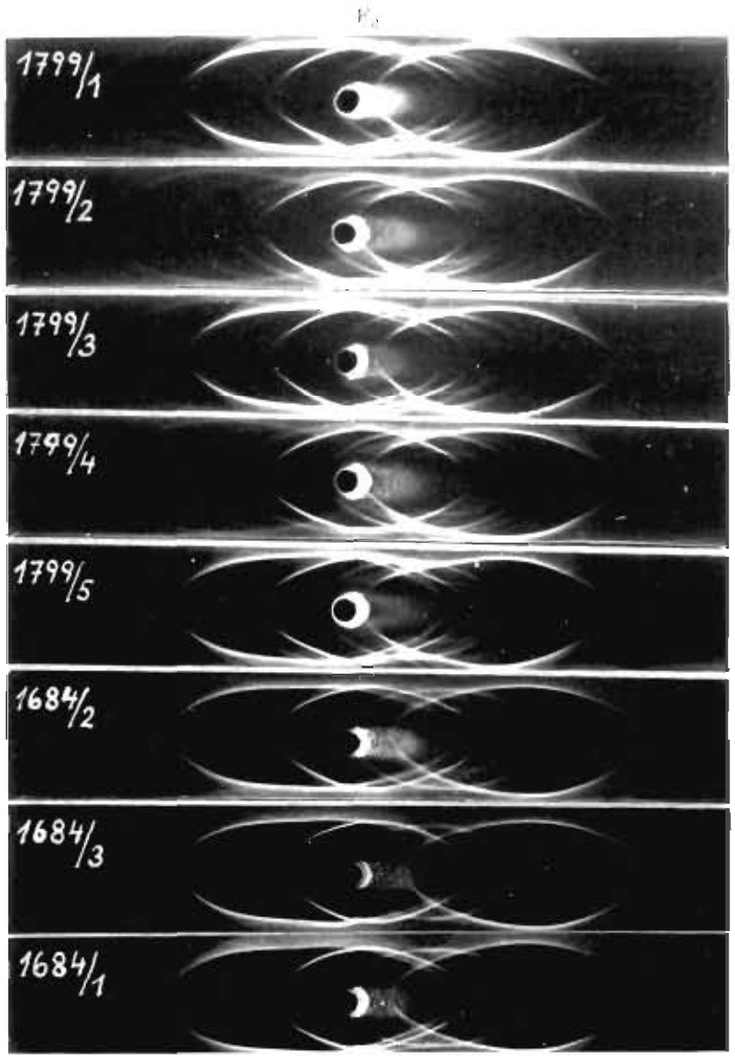


FIG. 8 (contd.). For legend see facing page.

case of polymorphism of protein II. In the cultivar Präsident Roosevelt this polymorphism was evidently removed in some samples by occasional selection from an originally heterogeneous population. Thus, polymorphism in protein II is more frequent and we can postulate that

such heterozygosity might be found frequently in land races and in cultivars obtained by means of mass selection.

Recently we have even found polymorphism in protein II between different seeds in a population of the cultivar Désirée of *P. coccineus* (see Fig. 7). This is not so surprising because *P. vulgaris* and *P. coccineus* are closely related species (or subspecies?) and gene introgression might be expected.

Finally, we have analysed protein II in intraspecific hybrids between types of *P. vulgaris* with (+) and without (−) this protein (see Fig. 8). In the F_1 generation [*P. vulgaris* cv. Veltruská Saxa (+) × *P. vulgaris* cv. Olympia (−)] the (+) type was dominant. In the F_2 there was segregation to both types in the ratio 564(+) types : 161(−) types, i.e. 3·50 : 1. This ratio agrees with the theoretical ratio 3 : 1, = 3·01, P = 20%, which shows that protein II is controlled by a single gene.

In summary it may be said that our results on the occurrence of two protein characters in *Phaseolus* (i.e. existence of polymorphism, loss mutation, transitional types, inheritance of two protein characters) can help in the evaluation of the taxonomic importance of single protein characters. They can also help in the selection of proteins in plant-breeding research.

11. Seed Protein Studies of *Brassica* and *Sinapis* Species

J. G. Vaughan

Department of Biology, Queen Elizabeth College,
University of London, England

ABSTRACT

The seed proteins of B. campestris, B. oleracea and B. nigra have been analyzed with serological and acrylamide gel electrophoresis techniques and the results are in good correlation with the established taxonomy. Further studies have been made of varietal status, generic status and amphidiploidy in *Brassica* and *Sinapis* using serological techniques and isoenzyme identification in starch gel.

The genus *Brassica* contains a number of species of great economic importance. These species are of interest to the plant taxonomist in terms of problems of speciation and hybridization (Bailey, 1930). In addition to studies of external morphology, data have also been accumulated concerning cytology (e.g. Moringa, 1934; U, 1935), comparative anatomy (e.g. Berggren, 1962) and volatile oil production (e.g. Vaughan *et al.*, 1963). Serological techniques, particularly those involving gel diffusion methods, have recently produced interesting results in connection with studies of angiosperm systematics (e.g. Gell *et al.*, 1960) and, consequently, it was felt that the application of these techniques to a study of *Brassica* might possibly further knowledge of the taxonomy of this difficult genus.

Although phytoserological methods are being more widely used by taxonomists, it was considered advisable in the first instance to apply these methods to a well accepted system of *Brassica* taxa (Vaughan *et al.*, 1966). Good correlation would support their further application to the more difficult problems of taxonomy. Three species were investigated in the first instance—*B. nigra* (Black Mustard), *B. oleracea* (Cabbage) and *B. campestris* (Turnip Rape). These species are well accepted as distinct taxa on morphological and cytological grounds.

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 93–102.

Antisera were raised in rabbits to saline extracts of the seed proteins of these plants and subsequently analyzed by the now well accepted double diffusion and immuno-electrophoretic methods. To effect a *complete serological comparison between the three species*, it was considered necessary to produce an antiserum for each taxon. Comparisons of the three species in double diffusion plates indicated specificity for

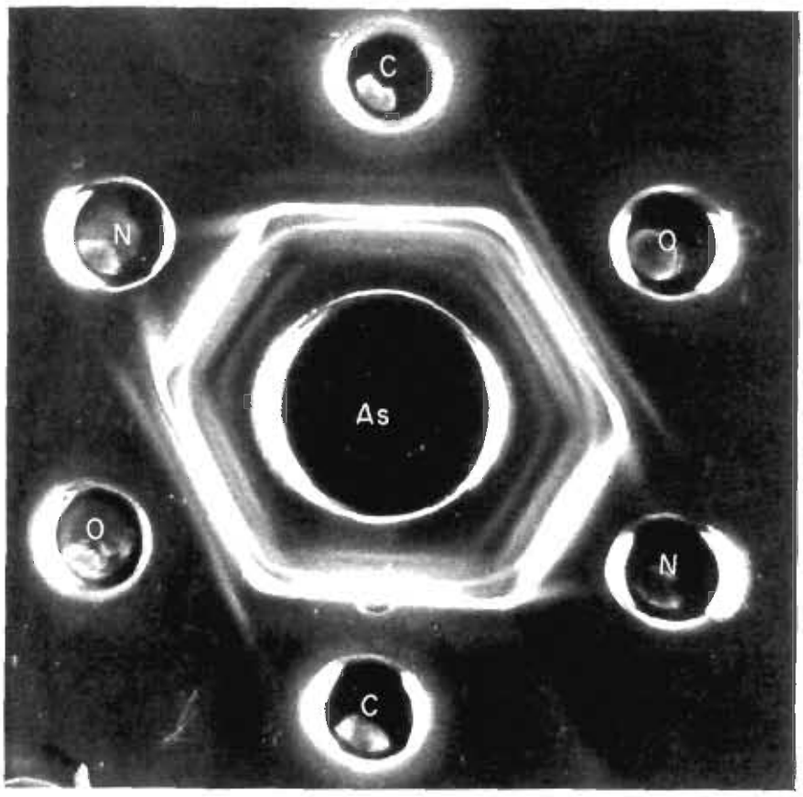


FIG. 1. Double diffusion plate. Antiserum (As) made to *Brassica oleracea* extract (O). Heterologous antigens are *B. campestris* (C) and *B. nigra* (N). Specific precipitin lines are shown against the homologous extract.

each taxon in terms of non-identical precipitin lines and spurs (Fig. 1) but the large number of antigen–antibody systems produced made a detailed comparison of the protein extracts very difficult. As expected, immuno-electrophoretic analysis gave a better resolution of the serological reactions but, with unabsorbed antisera, taxonomic comparison

was again difficult because of the large number of precipitin arcs. However, with all antisera produced, it was possible to distinguish *B. nigra* from *B. campestris* and *B. oleracea* in that it formed a strong arc towards the negative pole (Fig. 2, x). Similar strong arcs were formed by *B. campestris* and *B. oleracea* but either at the point of origin or slightly away from it.

To effect a more complete serological comparison, absorption of each antiserum was carried out with each of the heterologous extracts and the resulting antiserum analysed by double diffusion and immunoelectrophoresis against the homologous extract. The results of these experiments are presented in Table 1. Using differences in numbers of

TABLE 1

Antiserum (As) absorbed with heterologous antigen mixture (Ag)	No. of precipitin lines developing against homologous antigen mixture
$\frac{\text{AsC}}{\text{AgO}}$	5
$\frac{\text{AsC}}{\text{AgN}}$	9–10
$\frac{\text{AsO}}{\text{AgC}}$	1–2
$\frac{\text{AsO}}{\text{AgN}}$	4–6
$\frac{\text{AsN}}{\text{AgC}}$	3–5
$\frac{\text{AsN}}{\text{AgO}}$	3–5

C = *B. campestris*, O = *B. oleracea*, N = *B. nigra*

precipitin lines as a guide to taxonomic affinity, it would appear that *B. campestris* and *B. oleracea* are closer to each other than either is to *B. nigra*. This confirms the relationship based on the positions of the strong precipitin arcs obtained with the unabsorbed antisera.

In the same investigation (Vaughan *et al.*, 1966) a parallel study of the seed proteins of the three taxa was carried out by Boulter and Waiters using disc acrylamide gel electrophoresis. Their taxonomic results, based on a comparison of Rp values, confirmed those of the serological investigation. It would thus appear that the findings of this seed protein study are in agreement with taxonomy based on morphology and cytology not only in distinguishing the three species but

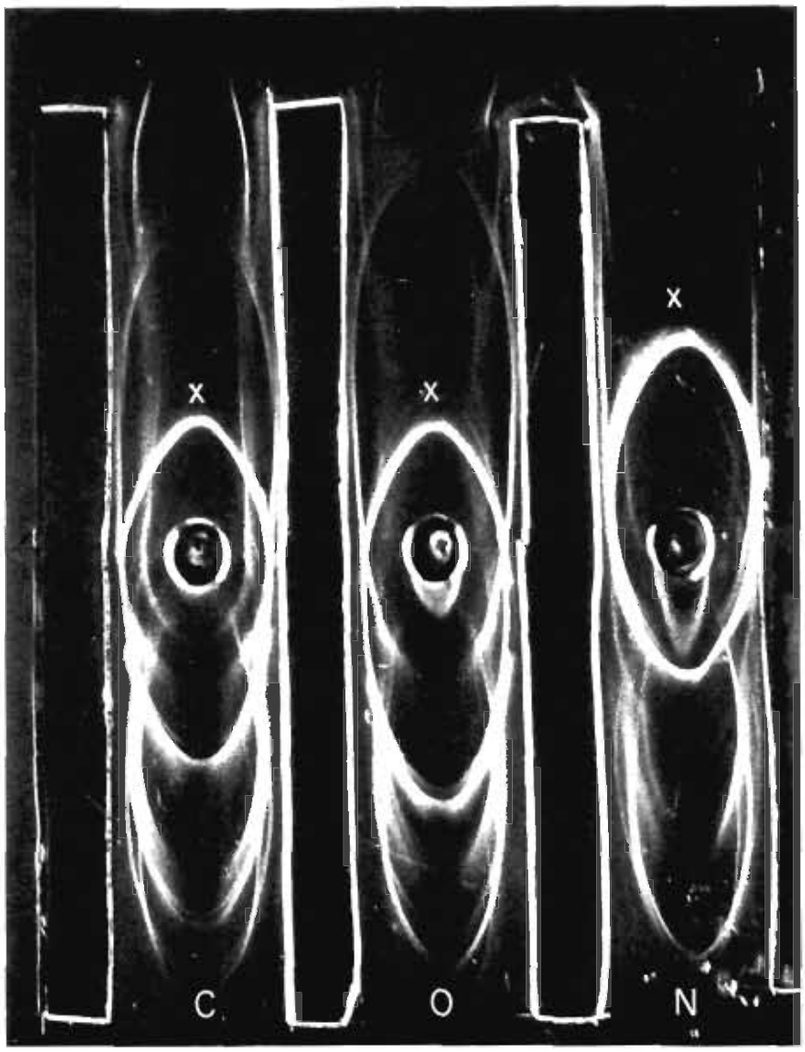


FIG. 2. Immuno-electrophoresis analysis of *Brassica campestris* (C), *B. oleracea* (O) and *B. nigra* (N) against *B. campestris* antiserum.

also in indicating a closer relationship between *B. campestris* and *B. oleracea* (Schulz, 1919).

As the result of this good correlation, the serological techniques described were next applied to more difficult problems of *Brassica*

taxonomy. The various forms of *B. campestris* have been the subject of much taxonomic discussion. Of these forms, turnip (*B. campestris* var. *rapa*) and turnip rape (*B. campestris* var. *autumnalis*) have sometimes been given varietal status, sometimes specific status (Musil, 1948). The seed proteins of these plants were analyzed serologically in conjunction with those of *B. oleracea* and *B. nigra* (Vaughan and Waite, 1967a; Table 2). In comparison with the other taxa, little serological difference

TABLE 2

Antiserum (As) absorbed with heterologous antigen mixture (Ag)	No. of precipitin lines developing against homologous antigen mixture	Antiserum (As) absorbed with heterologous antigen mixture (Ag)	No. of precipitin lines developing against homologous antigen mixture
As T / Ag C	2	As O / Ag T	2
As T / Ag O	5	As O / Ag C	1
As T / Ag N	8–9	As O / Ag N	5
As T / Ags C, O & N	1		
As C / Ag T	1 (very faint reaction)	As N / Ag T	3
As C / Ag O	4–6	As N / Ag C	3
As C / Ag N	7–8	As N / Ag O	4

T = *B. campestris* var. *rapa* C = *B. campestris* var. *autumnalis*
O = *B. oleracea* N = *B. nigra*

was demonstrated between turnip and turnip rape and consequently varietal status seemed to be indicated for these plants. Other serological relationships that appeared as the result of this work were that *B. nigra* seemed closer to *B. campestris* than to *B. oleracea* and that there was a closer relationship between turnip rape and *B. oleracea* than between turnip and *B. oleracea*.

The genus *Sinapis* has long been associated with *Brassica*. Linnaeus (1737) placed black mustard in *Sinapis* together with *S. alba* (white mustard). Schulz (1919) described black mustard as *B. nigra* but left white mustard in *Sinapis*. Bailey (1930) saw no reason for generic

distinction and included both forms in *Brassica*. Vaughan and Waite (1967a) compared serologically the seed proteins of *S. alba* with those of *B. nigra*, *B. campestris* and *B. oleracea* (Table 3). *S. alba* was closer to

TABLE 3

Antiserum (As) absorbed with heterologous antigen mixture (Ag)	No. of precipitin lines developing against homologous antigen mixture	Antiserum (As) absorbed with heterologous antigen mixture (Ag)	No. of precipitin lines developing against homologous antigen mixture
$\frac{\text{As S}}{\text{Ag C}}$	7	$\frac{\text{As O}}{\text{Ag S}}$	7
$\frac{\text{As S}}{\text{Ag O}}$	7	$\frac{\text{As O}}{\text{Ag C}}$	1
$\frac{\text{As S}}{\text{Ag N}}$	6	$\frac{\text{As O}}{\text{Ag N}}$	5–6
$\frac{\text{As C}}{\text{Ag S}}$	6	$\frac{\text{As N}}{\text{Ag S}}$	6
$\frac{\text{As C}}{\text{Ag O}}$	0	$\frac{\text{As N}}{\text{Ag C}}$	3
$\frac{\text{Ag C}}{\text{Ag N}}$	5	$\frac{\text{As N}}{\text{Ag O}}$	3

S = *Sinapis alba* C = *Brassica campestris* var. *autumnalis*
N = *B. nigra* O = *B. oleracea*

B. nigra than to *B. campestris* and *B. oleracea* but *B. nigra* was closer to *B. campestris* and *B. oleracea* than to *S. alba*. It is difficult to define the limits of a genus but these serological results suggest that any replacing of *B. nigra* also requires movements of *B. campestris* and *B. oleracea*.

These serological investigations were accompanied by starch gel electrophoresis studies of certain enzyme systems (esterases, β-glucosidases and β-galactosidases). No differences were revealed between turnip and turnip rape and, as this is in contrast to the enzyme variation shown between the other taxa, it would seem to support the serological indication of the varietal status of these plants. Apart from one esterase system shared with all the other taxa, *S. alba* had no esterases, β-glucosidases or β-galactosidases in common with the other taxa, support possibly for its generic status as indicated by the serological work. The distribution of esterases between *B. campestris*, *B. oleracea* and *B. nigra* indicated specificity but gave no information on relationships. However,

the distribution of β-glucosidases and β-galactosidases supported the serological and acrylamide gel results in demonstrating a close relationship between *B. campestris* and *B. oleracea* and a closer relationship between *B. nigra* and *B. campestris* than between it and *B. oleracea*.

Vaughan and Waite (1967b) investigated *B. carinata* (n = 17), *B. juncea* (n = 18) and *B. napus* (n = 19) which have been regarded as amphidiploid hybrids of *B. nigra* (n = 8), *B. oleracea* (n = 9) and *B. campestris* (n = 10). Seed esterases, β-glucosidases and β-galactosidases of all the taxa were compared by starch gel electrophoresis. The results strongly indicated that these enzyme systems of *B. carinata*, *B. juncea* and *B. napus* are essentially a summation of some of the proteins of the putative parents suggested by other botanical work, although a few specific proteins did appear.

The seed protein work described is being continued with special reference to the chemical identification of antigens of taxonomic importance and the protein changes occurring in species synthesis experiments.

ACKNOWLEDGMENT

I should like to thank the Agricultural Research Council for its support of this work.

REFERENCES

Bailey, L. H. (1930). The cultivated brassicas. *Gentes Herb.* **2**, 211–67.

Berggren, G. (1962). Reviews on the taxonomy of some species of the genus *Brassica* based on their seeds. *Svensk bot. Tidskr.* **56**, 65–135.

Gell, P. G. H., Hawkes, T. G. and Wright, S. T. C. (1960). The application of immunological methods to the taxonomy of species within the genus *Solanum*. *Proc. R. Soc.* Series B **151**, 364–83.

Linnaeus, C. (1737). "Genera Plantarum". Edition I. Lugdini Batavorum.

Moringa, T. (1934). Interspecific hybridization in *Brassica*. VI. The cytology of F1 hybrids of *B. juncea* and *B. nigra*. *Cytologia* **6**, 62–7.

Musil, A. F. (1948). Distinguishing the species of *Brassica* by their seed. *Misc. Publs U.S. Dep. Agric.* No. 643.

Schulz, O. E. (1919). Cruciferae–Brassiceae. *In* Engler's "Das Pflanzenreich", 1–290.

U, N. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jap. J. Bot.* **7**, 389–452.

Vaughan, J. G., Hemingway, J. S. and Schofield, H. J. (1963). Contributions to a study of variation in *Brassica juncea* Coss & Czern. *J. Linn. Soc.* (*Bot.*) **58**, 435–47.

Vaughan, J. G., Waite, A., Boulter, D. and Waiters, S. (1966). Comparative studies of the seed proteins of *Brassica campestris*, *Brassica oleracea* and *Brassica nigra*. *J. exp. Bot.* **17**, 332–43.

Vaughan, J. G. and Waite, A. (1967a). Comparative electrophoretic studies of the seed proteins of certain species of *Brassica* and *Sinapis*. *J. exp. Bot.* **18**, 100–109.

Vaughan, J. G. and Waite, A. (1967b). Comparative electrophoretic studies of the seed proteins of certain amphidiploid species of *Brassica*. *J. exp. Bot.* **18**, 269–76.

12. The Fluorescent Antibody Technique for the Identification of Plant Pathogenic Fungi

T. F. Preece

Agricultural Botany Division, School of Agricultural Sciences, University of Leeds, England

ABSTRACT

Serotaxonomic studies of fungi at Leeds include the production of specific rabbit gamma-globulins for the identification of fungi. Results obtained by labelling with fluorescein isothiocyanate are described. Such cross-reactions as do occur when the fluorescent antibody technique is used may be direct microscopic tools in taxonomy.

It has often been difficult to learn many details of plant disease epidemiology because of problems of distinguishing similar fungi. The mycelia of many, very different, fungi look similar when examined by direct light microscopy, and detailed cultural and biometric studies are needed for identification. Work at Leeds has concentrated on the utilization of the fluorescent antibody technique to determine specific fungi, without the need for culturing, on and in plants, and in soil.

Since the work of Coons and co-workers in the United States 20 years ago, when pneumococcal antigen in tissues was demonstrated by the use of an antiserum labelled with a fluorescent dye, this technique has been the subject of hundreds of papers concerned with a wide variety of topics: the fluorochromes themselves and their union or conjugation with proteins; the peculiar properties of conjugated serum proteins; the use of ultra-violet light microscopy to detect the union of labelled antibody with antigen, together with the immunological tracing of particular proteins, tissue antigens and antibodies, and of particular micro-organisms. The fluorescent antibody technique (F.A.T.) has been especially useful with the identification of bacteria. To a much lesser

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 111–114.

extent it has been utilized in the immunological tracing of protozoa, helminths, viruses, rickettsiae—and probably least of all, with fungi.

We have used, in the main, *fluorescein-iso-thiocyanate* (*Baltimore* Biological Laboratories F.I.T.C.) as our fluorochrome. Mycelium and spores of fungi may be detected by the corresponding fluorescein labelled antiserum.

The steps in the procedure for producing the tracer reagent, and its subsequent use are outlined in Table 1. Obligate plant pathogenic fungi

TABLE 1

Flow Diagram: Production and use of F.I.T.C. labelled Antibody for the Detection of Plant Pathogenic Fungi

PREPARATION OF ANTIGENS	Isolate fungus in pure culture Bulk up in liquid medium Filter off mycelium; grind.
PRODUCTION OF ANTIBODIES	Inoculate rabbit with formalized suspension of mycelium Detect extent of reaction with antigen by agar gel diffusion.
CONJUGATION WITH FLUORESCEIN-ISO-THIOCYANATE	Add dilute serum to F.I.T.C. in the cold Remove unreacted F.I.T.C. with charcoal Cross absorption may be necessary.
SPECIFIC STAINING	Fix slide culture of fungus with acetone Stain some preparations with F.I.T.C. conjugated antiserum Wash in buffered saline; dry.
MICROSCOPY	Mount slide in glycerol-saline Oil to condenser (U.V. inert immersion oil) Examine using U.V. light and dark-ground.
PHOTOMICROGRAPHY	35 mm film colour or black and white.

cannot be grown in culture. In these cases spores harvested from the surface of infected plants must be used to immunize the rabbit. Many important pathogens can be grown in artificial culture however. Once a suitable medium is found, preferably with a simple nitrogen source such as di-calcium glutamate, and glucose as a carbon source, the fungus is bulked up in liquid culture. Shaking the flasks has not helped greatly. The mycelium is filtered off with a Buchner funnel and the dry weight of an aliquot determined so that rabbits may be inoculated with a known dry weight of mycelium. A dense suspension of the mycelium is made by grinding a portion of wet mycelium. This is frequently difficult

and ultrasonic disintegration has not helped. Grinding by hand in a Griffith's tube often turns out to be the most satisfactory. Rabbits are injected with the formalized mycelial suspension both intravenously and subcutaneously, together with Freund's adjuvant, for a period of several weeks. When a suitable titre is obtained, or after about 4 weeks, the animal is bled by cardiac puncture. On occasion low titres have given quite sensitive reagents after conjugation with F.I.T.C. and a high titre is not, it seems, essential. The titre is most conveniently obtained by the use of the Ouchterlony agar gel diffusion plate method. At first only gamma-globulin fractions of rabbit serum were used, but tests showed that these were no better as diagnostic reagents than conjugated whole sera, and these are in use for several pathogens now. In the early work, the protein concentration of the serum was determined by a modification of the Folin and Ciocalteau colorimetric method so that an amount of F.I.T.C. could be added which was proportional to the protein content. Now we add an excess of F.I.T.C., *c.* 1 mg/ml and remove the excess, after conjugation, by charcoal absorption. The charcoal process leads inevitably to slight loss of antibody content and this may be reduced by using as little activated charcoal as possible, less than 150 mg/ml of original serum, and better still, by the use of a Sephadex column.

Bacteriologists are easily able to heat fix their organisms on a slide. This is not applicable to fungi. Considerable distortion results from the heating in a Bunsen flame, and more important still, the hyphae are difficult to locate and easily washed off. Techniques developed here by Mrs. Dorothy Cooper have surmounted this difficulty by the use of special micro-slide cultures. The mycelium, when grown on a slide, is very difficult to remove, and stains well with the F.I.T.C. conjugated antiserum.

The slide is fixed in acetone and the F.I.T.C. labelled antiserum left on the slide for 30 min. This is then washed off by two changes of buffered saline, followed by a rinse in distilled water. The dry preparation is covered with glycerol saline beneath a cover slip and examined by darkground microscopy using a Vickers H.B. 200 mercury lamp in a dark room. Control specimens of other fungi are used, and the unstained fungus itself checked for autofluorescence. Where necessary, cross-reacting antibodies to antigens of other fungi are removed by absorption.

Photomicrography in colour of the apple green stained hyphae on High Speed Ektachrome usually requires an exposure of 10 min. The best black and white results have been obtained by using HP3 Ilford film, exposures being about 1 min. Typical photomicrographs show, for example, the mycelium of *Botrytis cinerea* stained with an F.I.T.C.

conjugated antiserum, on HP3 Ilford film, exposure 50 sec, magnification ×250. Unstained mycelium of the fungus was invisible under U.V. microscopy. This result is quite typical.

Detecting the fungus in tissue has other problems, notably the autofluorescence of plant tissue. The use of coloured counterstains seems likely to overcome this difficulty.

SECTION III

LARGE MOLECULES (NUCLEIC ACIDS)

13. Taxonomic Applications of DNA Hybridization Techniques

D. E. KOHNE

Department of Terrestrial Magnetism, Carnegie Institution, Washington, D.C., U.S.A.

ABSTRACT

DNA hybridization provides a means for detecting the DNA sequences held in common between two species. The experimental technique for these studies depends upon separating unreassociated DNA (single strand) from DNA which has reassociated (double strand). The properties of calcium phosphate crystals (hydroxyapatite) provide a simple tool for separating single-strand from double-strand DNA. Several DNA comparison experiments illustrate the usefulness of this technique.

DNA from different life forms varies in its complexity. Bacterial and viral DNAs are composed almost entirely of DNA sequences which are present only once per virus particle or bacterial cell. A major portion of higher organism DNA, however, is composed of DNA sequences which are repeated a large number of times. The implications of this complexity to existing DNA relationship data are discussed.

Methods have recently been developed in our laboratory which allow accurate and reproducible DNA comparisons among different life forms (Britten and Kohne, 1967). Since these techniques utilize physical properties of the nucleic acid molecules it will be helpful to describe briefly the structure and properties of the DNA molecule which allow these DNA sequence comparisons to be made.

DNA is in the form of long double-stranded molecules. These molecules are termed "native" as long as the two strands have not been separated *in vitro*. The single strands of the DNA molecule are composed of long linear sequences of the four bases A, G, C, T (adenine, guanine, cytosine and thymine). These sequences contain information written in the language of the genetic code.

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 117–130.

The two single strands of the native DNA molecule are held together principally by specific interactions between complementary nucleotide pairs. A and T pair only with each other just as G and C only interact together. A and T, and G and C are referred to as complementary base or nucleotide pairs.

The DNA molecule then, consists of two polynucleotide strands, arranged in a helical configuration, in which the bases of one strand are joined with complementary bases in the second strand (Marmur *et al.*, 1963). The two strands in a double-stranded DNA molecule are referred to as complementary strands.

The two complementary strands of the double-stranded DNA molecule may be completely separated (dissociated) from each other (Fig. 1). Under the proper conditions the dissociated complementary

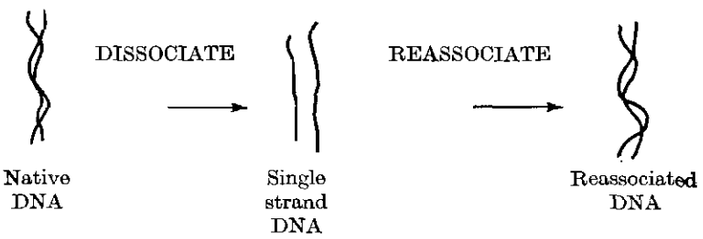


FIG. 1. A schematic view of dissociation and reassociation of DNA.

strands will *reassociate* and form stable double-stranded molecules with a helical structure similar to that of native DNA.

Since the reassociation reaction involves the collision of two complementary strands the *rate* of the reaction is dependent upon the concentration of appropriate DNA strands. At high concentration of a

---

A High Concentration of Non-radioactive DNA from Species A. This DNA Will Reassociate Rapidly and Completely.

+

A Very Low Concentration of Radioactive DNA from Species B. This DNA will Reassociate Only if There are Similar DNA Sequences in the Non-radioactive DNA of Species A.

↓

Dissociate and Incubate Under Reassociating Conditions

↓

Separate Reassociated DNA from Dissociated (single-stranded) DNA

↓

The Amount of Radioactivity Present in the Reassociated Fraction is a Measure of the Extent of DNA Relatedness Between the Two Species

FIG. 2. The basic experimental procedure used for DNA comparison studies.

DNA (and therefore a high concentration of each complementary sequence) the reassociation reaction is relatively rapid. At very low concentrations of DNA the reaction is very slow.

The reassociation reaction provides a means for detecting the DNA sequences held in common between two species. The technique is simple and straightforward (Fig. 2). A small amount of radioactive DNA from species B is mixed with a large amount of non-radioactive DNA from species A. The DNA in the mixture is dissociated and then placed under conditions where the complementary strands may reassociate. The concentration of the radioactive DNA is made very low in order that it cannot reassociate with itself. A strand of radioactive DNA can under these conditions only reassociate if it has a complementary strand in the non-radioactive DNA from species A. The con-

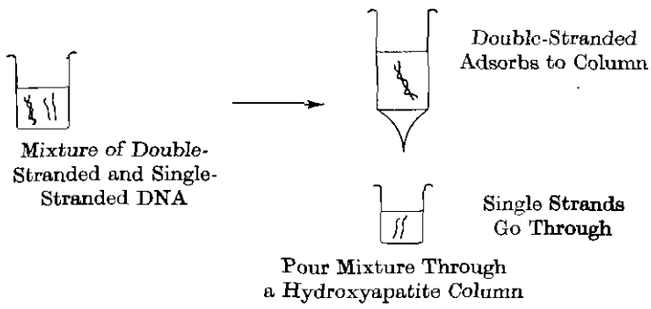


FIG. 3. A schematized view of the use of a hydroxyapatite column to separate single-stranded DNA from double-stranded DNA.

centration of the species A DNA is made very high and therefore reacts rapidly and completely.

At the end of the reaction the reassociated DNA is separated from the non-reassociated DNA. Since the radioactive DNA could not react with itself the amount of radioactivity in the reassociated fraction is a measure of the DNA sequences held in common between the two species, A and B.

A simple technique has been developed to separate reassociated double-stranded DNA from dissociated single-stranded DNA (Britten and Kohne, 1967). Crystals of calcium phosphate (hydroxyapatite) have the ability to discriminate between double-stranded and single-stranded DNA. Under the proper salt and temperature conditions hydroxyapatite will adsorb double-stranded DNA but not single-stranded DNA (Fig. 3). To separate reassociated double-stranded

DNA from single-strand DNA the DNA reassociation mixture is passed through a hydroxyapatite column. The non-reassociated DNA (i.e. single-stranded DNA) passes through the column while the re-associated DNA adsorbs on to the hydroxyapatite. The reassociated DNA is readily recovered from the column by either salt or temperature elution.

The usefulness of the technique just described can be illustrated further by considering an experiment carried out by Brenner and Cowie (unpublished). This experiment was designed to ask two questions: 1. What is the extent of DNA sequence relationship detectable between the DNAs of several species of bacteria? 2. When relatedness is detected between two bacterial DNAs are the interacting strands completely complementary to each other?

For this experiment a small amount of radioactive DNA was mixed with a large amount of non-radioactive DNA from other bacteria. These mixtures were then dissociated and incubated for an appropriate time. At the end of the incubation the mixtures were passed over hydroxyapatite to isolate the reassociated fractions. Table I presents the measured relatedness observed between *Escherichia coli* and several other bacteria.

We have seen in Table 1 how much of the *E. coli* DNA is held in

TABLE 1. Observed relatedness of *E. coli* DNA to the DNAs of different bacterial species. The percent relatedness is the ratio of the fraction of the labelled *E. coli* DNA which reassociates with the DNA of another species over the fraction of labelled *E. coli* DNA which reacts with unlabelled *E. coli* DNA × 100. The re-association conditions for all samples were comparable.

BACTERIA	Relatedness to *E. coli*
E. coli	100%
Shigella flexneri	85%
Salmonella typhimurium	35%
Aerobacter aerogenes	35%
Proteus mirabilis	2–5%

common with other bacteria. The next question to be asked was: Are the DNA sequences which are held in common between these bacteria really identical?

It is possible to measure the degree of similarity between the DNA strands of helical double-stranded DNA (reassociated or native) (Britten and Kohne, 1967). The thermal stability (i.e. the temperature at which the double strands separate into two single strands) of perfectly base pair matched double-stranded DNA is higher than that

of double-stranded DNA which has some degree of base pair mismatches. The greater the degree of base pair mismatching the lower the thermal stability of the reassociated DNA. Native DNA is, of course, assumed to have perfect base pair matching.

In bacteria the intraspecies reassociated DNA has a thermal stability which is virtually identical to that of native bacterial DNA of that species. It is possible to determine the precision of base pair matching of the interspecies reassociated DNA (e.g. one strand of the reassociated double-stranded molecule came from *E. coli* and one from *Salmonella*) by comparing its thermal stability to that of the reassociated *E. coli* DNA.

Hydroxyapatite is also a useful tool for measuring the thermal stability of double-stranded DNA. Double-stranded DNA binds to hydroxyapatite while single-stranded DNA passes through the column. If the temperature of the hydroxyapatite is raised above the temperature at which the double strands separate, the single-stranded DNA can be washed off the hydroxyapatite column and recovered and measured. A *thermal elution profile* for a double-stranded DNA can be obtained by raising the temperature of the hydroxyapatite column in successive steps and washing away all of the DNA rendered single-stranded after each temperature rise (Britten and Kohne, 1967).

Figure 4 presents the thermal elution profiles of reassociated *E. coli* DNA and the reassociated double-stranded molecules composed of both *E. coli* and *Salmonella* DNA strands. The reassociated *E. coli* DNA has a much higher thermal stability than does the interspecies double-stranded hybrid. This indicates that the base pair matching in the interspecies hybrid is less than perfect. The detectable sequences held in common between *E. coli* and *Salmonella* are then, similar enough to react together to form a stable product but are not identical. Therefore extensive changes in the DNA of *E. coli* and *Salmonella* have occurred since their divergence from one another.

Comparison of DNAs from different higher organisms can also be accomplished by using the techniques already described. DNA from higher organisms is, however, more complex than bacterial DNAs in two ways. 1. Higher organisms coutaiu much more DNA per cell (McCarthy, 1965). 2. Many individual DNA sequences in higher organism DNA are repeated a very large number of times with some sequences being repeated as much as one million times while other DNA sequences in the same genome are unique.

The repeated DNA sequences are readily detected by their rapid rate of reassociation relative to the rate of reassociation of the DNA which is not repeated. Two complementary strands must collide in order to reassociate. DNA sequences which occur more frequently than other

DNA sequences will collide with complementary strands more often and will therefore reassociate more rapidly than the less frequent sequences. For example, if a sequence is present 100 000 times per cell it will reassociate 100 000 times more rapidly than a sequence which is present only once per cell.

The hydroxyapatite method also provides a tool with which to separate rapidly reassociating DNA from slowly reassociating DNA.

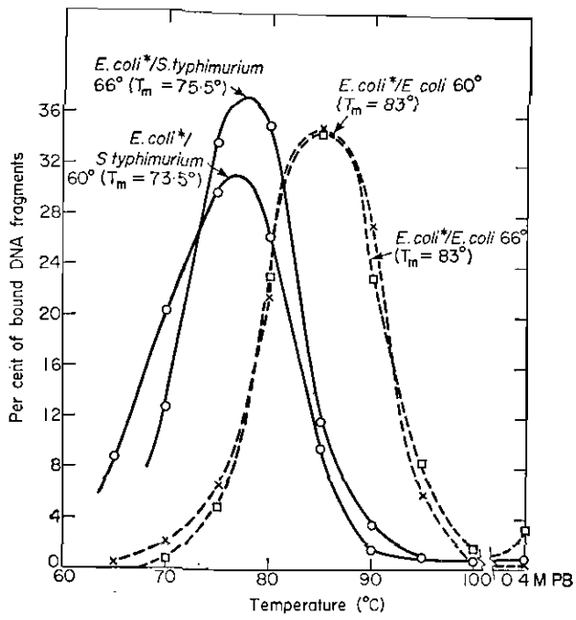


FIG. 4. Thermal elution profiles of reassociated bacterial DNAs from hydroxyapatite. Broken lines: thermal elution profile of reassociated DNA composed of labelled *E. coli* DNA and unlabelled *E. coli* DNA. Solid lines: thermal elution profiles of reassociated DNA composed of one strand of labelled *E. coli* DNA and one strand of unlabelled *Salmonella* DNA. Incubations were performed at two temperatures, 60° C and 66° C.

The technique simply involves: (a) dissociating the DNA; (b) reassociating the DNA for an appropriate time; (c) passing the reassociation mixture through hydroxyapatite where the rapidly reassociating DNA (which has reassociated) is adsorbed while the single-stranded slowly reassociating DNA passes through the column. These isolated DNA fractions can then be further characterized and fractions from different organisms can be compared and characterized.

The DNAs of a variety of creatures have been examined in order to

determine the distribution of repetitious DNA among life forms. The species of bacteria and blue green algae (*E. coli, Clostridium perfringens, Plectonema boryanum*) that have been examined do not contain repetitious DNA detectable by these methods (Britten and Kohne, 1968).

The list of higher organisms containing repetitious DNA is presented in Table 2. All organisms "higher" than bacteria contain repetitious

TABLE 2. A list of higher organisms whose genomes contain sizeable repeated DNA fractions. All higher organisms checked thus far contain repeated DNA.

PROTOZOA
Dinoflagellate (*Gyrodinium cohnii*)[a]
Euglena gracilis[a]
PORIFERA
Sponge (*Microciona*)[a]
COELENTERATES
Sea anemone (*Metridium*) (tentacles)[a]
ECHINODERMS
Sea urchin (*Strongylocentrotus*) (sperm) [a, b, c]
Sea urchin (*Arbacia*) (sperm)[a, b, c]
Starfish (*Asterias*) (gonads)[a]
Sand dollar (*Echinarachnis*)[c]
ARTHROPODS
Crab (*Cancer borealis*) (gonads)[a]
Horseshoe crab (*Limulus*) (hepatopancreas)[a]
MOLLUSCS
Squid (*Loligo pealii*) (sperm)[a]
ELASMOBRANCHS
Dogfish shark (liver)[a]
OSTEICHTHYES
Salmon (sperm)[a, b, c]
Lungfish[a, c]
AMPHIBIA
Amphiuma (liver, red blood cells, muscle)[a]
Frog (*Rana pipiens*)[b]
Frog (*Rana sylvatica*)[c]
Toad (*Xenopus laevis*) (heart, liver, red blood cells)
Axolotl (*Ambystoma tigrinum*)[c]
Salamander (*Triturus viridescens*)[c]

TABLE 2—continued

AVES
Chicken (liver, blood)[a, b, c]
MAMMALS
Tree shrew[c]
Armadillo[c]
Hedgehog[c]
Guinea-pig[c]
Rabbit[c]
Rat (liver)[a, b, c]
Mouse (liver, brain, thymus, spleen, kidney)[a, b, c]
Hamster[c]
Calf (thymus, liver, kidney)[a, b, c]
PRIMATES
Tarsier[c]
Slow Loris[c]
Potto[c]
Capuchin[c]
Galago[c]
Vervet[c]
Owl monkey[c]
Green monkey[c]
Gibbon[c]
Rhesus[b, c]
Baboon[c]
Chimp[a, c]
Human[a, b, c]
PLANTS
Rye (*Secale*)[c]
Tobacco (*Nicotiana glauca*)[c]
Bean (*Phaseolus vulgaris*)[c]
Vetch (*Vicia villosa*)[c]
Barley (*Hordeum vulgare*)[a, b]
Pea (*Pisum sativum* var. Alaska)[a, b]
Wheat (*Triticum aestivum*)[a, c]
Onion (*Allium* sp.)[a]

[a] Rate of reassociation measured directly by hydroxyapatite fractionation and/or measurement of optical hypochromicity as a function of time.

[b] Labelled, sheared fragments bind to DNA from the same species imbedded in agar.

[c] Sheared non-radioactive fragments of DNA from the listed organism compete with the DNA-agar reaction (*b*) of a related species, reducing the amount of labelled DNA which binds to the imbedded DNA.

DNA. Virtually all of onion and salmon DNA is repetitious while repeated DNA comprises about 30–50% of mouse and cow DNA. The fraction of the DNA observed as repeated DNA varies with the species. The pattern of repetition of DNA sequences of mouse DNA is presented in Fig. 5. This shows the amount of DNA which occurs at different repetition frequencies. One fraction of about 10% of the DNA is composed of a DNA sequence which is repeated about *one million* times. About 15% of the DNA has an average repetition frequency ranging from 1 000 to 100 000 while the majority of the DNA is composed of unique sequences.

Early studies on DNA comparisons were carried out by Bolton, Hoyer and McCarthy (1964) among creatures whose taxonomic relationships are well established. The method they used (Hoyer *et al.*,

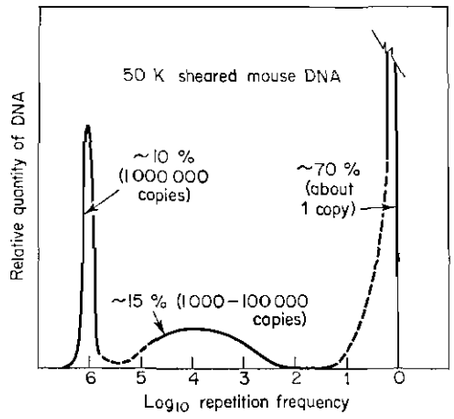


FIG. 5. A repetition frequency spectrogram for mouse DNA. The relative quantity of DNA is plotted against the logarithm of the repetition frequency. The broken segments of the curve represent regions of some uncertainty. The evidence strongly suggests that the peak at the right is composed primarily of unrepeated DNA sequences (Britten and Kohne, 1967). These data were derived from the measurements of the quantity and rate of reassociation of fractions separated on hydroxyapatite.

1964) involves immobilizing long DNA strands from one animal in an agar matrix and reacting this trapped DNA with shorter strands of DNA from another source. At the end of the reaction the shorter strands of DNA which did not react with the trapped DNA are washed away and the amount of labelled DNA adsorbed to the DNA-agar measured (Fig. 6).

Table 3 presents data obtained from a series of these reactions in

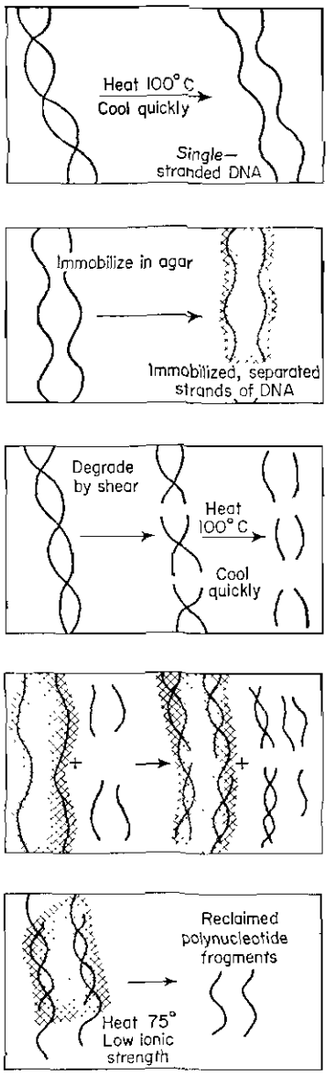


FIG. 6. Schematic representation of the *DNA-agar procedure.*

which immobilized mouse DNA was reacted with short radioactively labelled DNA strands from different vertebrates.

Labelled mouse DNA fragments react with mouse DNA-agar to an extent of about 30%. This 30% fraction represents the rapidly re-associating DNA of the mouse. This result illustrates the fact that only

TABLE 3. Observed relatedness of mouse DNA-agar with labelled DNAs from other vertebrate species (see Fig. 7).

Animal DNA	% Relatedness to Mouse
RAT	74
HAMSTER	60
MAN	35
RHESUS	29
BOVINE	27
CHICKEN	6

the repetitious DNA of higher organisms can reassociate under the conditions used for the DNA-agar technique. The data presented in Table 3 then represent the degree of relatedness observed among the *repeated* DNA sequences of various vertebrates.

Figure 7 presents these same data in a different way and shows the

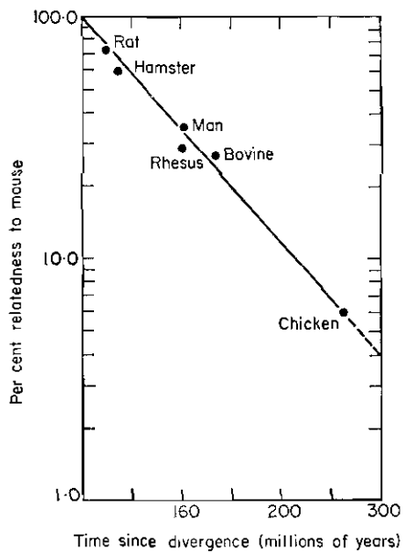


FIG. 7. The reaction of different vertebrate DNA fragments with mouse DNA-agar. The percent relatedness given here is the ratio of the extent of interspecies binding to the extent of intraspecies binding (mouse on mouse) $\times 100$. Values for the time since divergence of the various animals were obtained from the palaeontological literature.

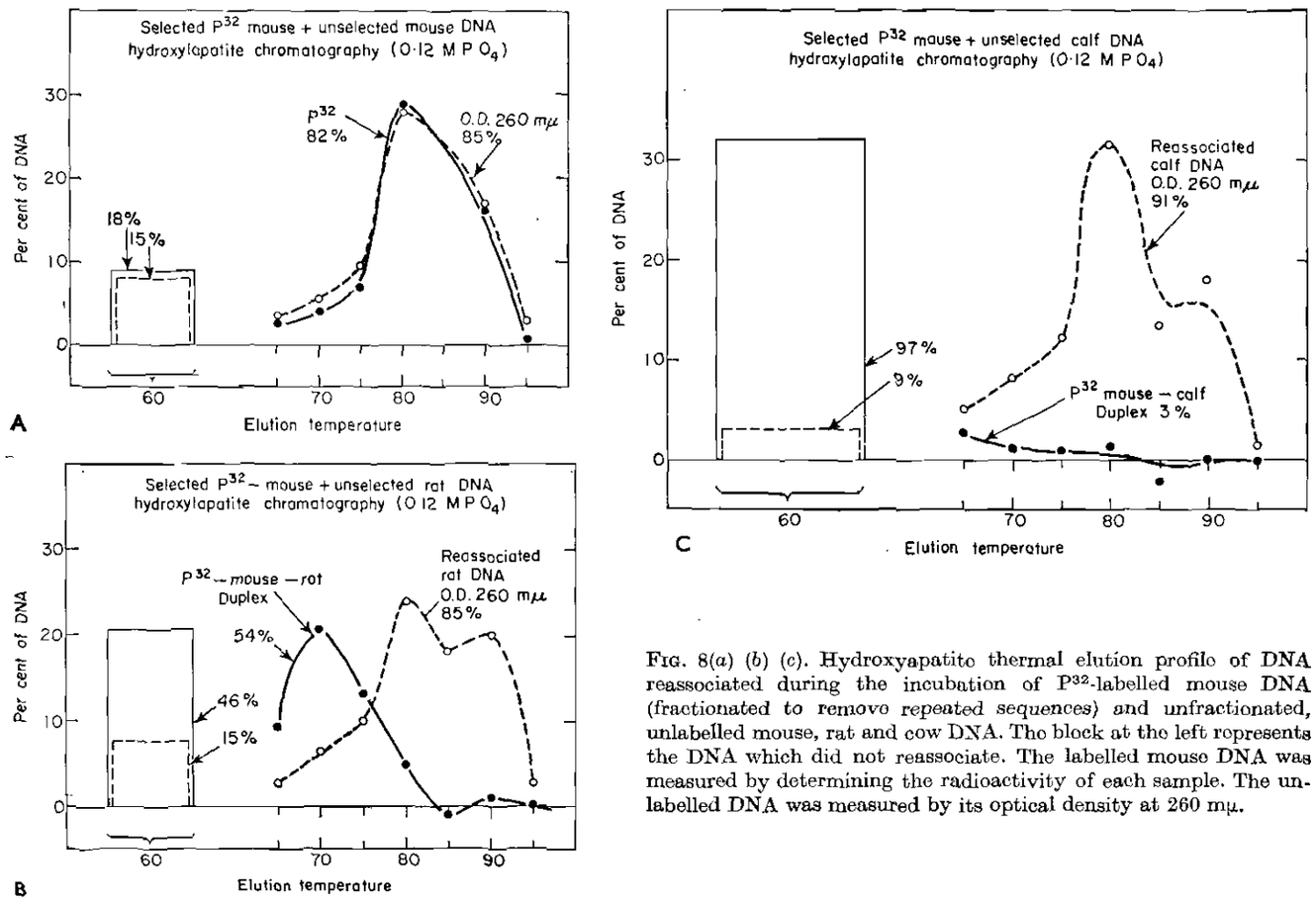


FIG. 8(*a*) (*b*) (*c*). Hydroxyapatite thermal elution profile of DNA reassociated during the incubation of P^{32}-labelled mouse DNA (fractionated to remove *repeated sequences*) and unfractionated, unlabelled mouse, rat and cow DNA. The block at the left represents the DNA which did not reassociate. The labelled mouse DNA was measured by determining the radioactivity of each sample. The unlabelled DNA was measured by its optical density at 260 mμ.

observed loss of relationship among the repetitious DNA fractions with evolutionary time. On this graph there appears to be a linear correlation between the degree of DNA relatedness observed and the time since divergence of the different vertebrates. The straight-line aspect of this graph has obvious systematic implications and further work of this type is in order.

When it became possible to fractionate higher organism DNA and isolate and purify the slowly reassociating fraction of a higher organism DNA it was of great interest to compare the slowly reassociating DNA fractions of various animals. For these experiments isolated labelled slowly reassociating DNA from the mouse was reacted with unfractionated DNA from the cow and rat. After an appropriate incubation time the reaction mixtures were passed over hydroxyapatite and thermal elution profiles of the reassociated DNA were obtained. Figure 8 (*a*, *b*, *c*) present data obtained from these experiments. About 60% of the labelled mouse slow fraction DNA is able to react with rat DNA. The precision of base pair matching between the sequences held in common between mouse and rat however is very poor. Little if any mouse slow fraction DNA is able to reassociate with cow DNA. These results indicate that very extensive sequence changes have occurred in mouse and cow DNA since their divergence while, as expected, mouse and rat DNAs are much more similar.

Relatively little has been done towards utilizing DNA comparisons as a tool for taxonomic and systematic studies. The development of the hydroxyapatite technique for measuring DNA hybridization now provides a simple, powerful tool with which to attack these problems. In addition to the work reported here DNA comparisons have been done for viral (Cowie and Szafranski, 1967), bacterial (De Ley and Park, 1966) and plant (Bendich and Bolton, 1967) DNAs. The results of all these experiments further suggest the immense potential of this technique as a tool for taxonomy and systematics.

REFERENCES

Bendich, A. J. and Bolton, E. T. (1967). Relatedness among plants as measured by the DNA-agar technique. *Pl. Physiol. Lancaster* **42**, 959–67.

Bolton, E. T., Hoyer, B. H. and McCarthy, B. J. (1964). Biophysics Report. *Carnegie Inst. Wash. Yr. Bk.* **63**, 366–97. Garamond/Pridemark Press, Baltimore, Maryland.

Britten, R. J. and Kohne, D. E. (1967). Biophysics Report. *Carnegie Inst. Wash. Yr. Bk.* **65**, 78–123.

Britten, R. J. and Kohne, D. E. (1968). Biophysics Report. *Carnegie Inst. Wash. Yr. Bk.* **66**, 66–86.

Cowie, D. B. and Szafranski, P. (1967). Thermal chromatography of DNA–DNA reactions. *Biophys. J.* **7**, 567–84.

De Ley, J. and Park, I. W. (1966). Molecular biological taxonomy of some free-living nitrogen-fixing bacteria. *Antonie van Leewenhoek* **32**, 6–16.

Hoyer, B. H., McCarthy, B. J. and Bolton, E. T. (1964). A molecular approach in the systematics of higher organisms. *Science N.Y.* **144**, 959–67.

Marmur, J., Rownd, R. and Schildkraut, C. L. (1963). Denaturation and renaturation of deoxyribonucleic acid. *Prog. Nucl. Acid Res.* **1**, 231–300. Academic Press, New York.

McCarthy, B. J. (1965). The evolution of base sequences in polynucleotides. *Prog. Nucl. Acid Res.* **4**, 129–57. Academic Press, New York.

14. Evolution of Ribosomal RNA

HANNAH J. GOULD

M.R.C. Biophysics Research Unit, King's College, University of London, England

ABSTRACT

Past studies have shown that ribosomal RNA molecules from different sources display differences in size, composition, secondary structure, and mode of biosynthesis. Ribosomal RNA, together with the associated proteins which make up ribosomes, occurs in all living organisms. Thus it should be possible to trace the development of these molecules over the widest possible evolutionary span. This type of study ultimately necessitates the determination of the complete nucleotide sequence of ribosomal RNA from a range of organisms. The determination of the complete nucleotide sequence is a huge task, at present out of reach. A sensitive new method described here, however, may reveal whether two species of RNA have the same sequence or not. The method involves limited digestion of the RNA with nucleases and separation of the fragment by polyacrylamide gel electrophoresis. The molecular weight range for which polyacrylamide gels are applicable and the factors which determine the degree of resolution are discussed together with results on ribosomal RNAs isolated from bacteria, plant and animal species. Preliminary results on the characterization of ribosomal RNA fragments after separation on polyacrylamide gels are presented. These results lead to a number of generalizations about the nature of evolutionary changes in ribosomal RNA.

Ribosomes catalyze protein synthesis by the same basic mechanism in all cells. It is interesting therefore that ribosomes from various sources display structural differences. I wish here to examine the nature of these differences and speculate on their evolution. I will focus attention only on the RNA components of the ribosome with which my own work has been primarily concerned. In any case, studies on the proteins and other constituents of ribosomes have not been carried out on a sufficiently wide variety of organisms to permit general conclusions.

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 131–156.

After reviewing the earlier work bearing on the subject, I will describe some of our recent results using a new technique for characterizing and comparing RNA molecules. This technique involves limited digestion of RNA by nucleases and separation of fragments by polyacrylamide gel electrophoresis. Hence it is similar in principle to the "fingerprinting" of proteins (Ingram, 1958) which, owing to its contribution to amino acid sequence studies, has helped to place the science of taxonomy on a molecular basis.

PROPERTIES OF DIFFERENT TYPES OF RIBOSOMES

It has been recognized for some time that there are at least three distinct classes of ribosomes, differing in size and in the relative proportions of RNA and protein (Petermann, 1964; Taylor and Storck, 1964; Taylor *et al.*, 1967; Stutz and Noll, 1967). Bacterial, mitochondrial and chloroplast ribosomes together form one class distinguished by their relatively small size (Table 1). As Küntzell and Noll

TABLE 1. *Properties of ribosomes.*

Source	$S^{\circ}_{20,w}$[a]	MW × 10^{-6}	% RNA[a]
E. coli	70	3·1[b]	63
Yeast	80	4·1[c]	40
Pea seedlings	80	4·1[d]	40
Rabbit reticulocyto	80	4·1[e]	40
Rat liver	83	5·0[f]	45

[a] Petermann (1964)
[b] Tissières *et al.* (1959)
[c] Chao and Schachman (1956)
[d] Ts'o *et al.* (1958)
[e] Dintzis *et al.* (1958)
[f] Tashiro and Yphantis (1965)

(1967) have demonstrated there is a gradation in size, indicated by the sedimentation coefficients, $S^{\circ}_{20,w}$, even within the class of bacterial ribosomes, ranging from what is considered to be the most primitive organism studied, a photosynthetic bacterium *Rhodospirillum rubrum* ($S^{\circ}_{20,w}$ = 66), to the more advanced species *Bacillus licheniformis* ($S^{\circ}_{20,w}$ = 73).

Ribosomes contain two ribonucleoprotein subunits each with a single RNA chain. The size of the RNA in the larger subunit is approximately 2·5 times that of the smaller. It is of interest to note that the ribosomal RNAs of animal and plant cells, like the ribosomes themselves, are larger by 20–40% than the corresponding species in bacteria (Table 2).

TABLE 2. Molecular weights of ribosomal RNA.

Source	Molecular weight $\times 10^6$	
	Large component	Small component
E. coli[a]	1·1	0·5
Yeast[b]	1·4	0·7
Rat liver[c]	1·7	0·6

[a] Kurland (1960)
[b] Maeda (1961)
[c] Gierer (1958)

Significant variations in the base composition of ribosomal RNAs have also been observed. A general survey (Petermann, 1964) reveals that the larger RNA component of animal cell ribosomes contains a higher proportion of both guanine and cytosine than that of bacteria and yeast ribosomes (Table 3). An exception to the rule is the ribosomal RNA of *Drosophila melanogaster*, which has an unusually low G+C content (Hastings and Kirby, 1966) (Table 3).

TABLE 3. Base compositions of ribosomal RNA.

Base	Base composition							
	Large component				Small component			
	E. coli[a]	Yeast[b]	Rabbit reticulocyte[c]	*Drosophila melanogaster*[d]	*E. coli*[a]	Yeast[b]	Rabbit reticulocyte[c]	*Drosophila melanogaster*[d]
A	0·26	0·27	0·17	0·31	0·25	0·27	0·22	0·29
U	0·22	0·26	0·16	0·27	0·21	0·29	0·21	0·27
G	0·31	0·27	0·37	0·23	0·32	0·24	0·32	0·24
C	0·21	0·20	0·31	0·20	0·23	0·20	0·24	0·20
G + C	0·52	0·47	0·68	0·43	0·55	0·44	0·56	0·44

[a] Spahr and Tissières (1959)
[b] Spencer *et al.* (personal communication)
[c] Gould *et al.* (1966)
[d] Hastings and Kirby (1966)

Ribosomal RNA possesses segments of double helix in which the bases are paired as in DNA (Spirin, 1963; Spencer *et al.*, 1962; Spencer and Poole, 1965). The stability and composition of these regions may also be compared (Table 4). The guanine plus cytosine (G + C) content in the double helical region, like that of the overall base composition, increases in the order yeast < bacteria < rabbit reticulocyte. Concomitant with this, since GC base pairs are more stable than AU

pairs, we find the midpoint of the thermal transition from the paired to the unpaired state of the bases in 0·01 M phosphate buffer for *E. coli* ribosomal RNA is 46°C compared to 54°C for rabbit reticulocyte ribosomal RNA (Cox, 1966).

TABLE 4. Base compositions of paired and unpaired regions of ribosomal RNA.

Base	Base composition								
	E. coli[a]			Yeast[b]			Rabbit reticulocyte[a]		
	Paired	Unpaired	Total	Paired	Unpaired	Total	Paired	Unpaired	Total
A	0·22	0·38	0·27	0·15	0·11	0·26	0·17	0·19	0·18
U	0·22	0·16	0·20	0·15	0·11	0·26	0·17	0·21	0·18
G	0·28	0·33	0·29	0·14	0·14	0·28	0·33	0·43	0·36
C	0·28	0·12	0·23	0·14	0·05	0·19	0·33	0·17	0·28
G + C	0·56	0·45	0·52	0·28	0·19	0·47	0·66	0·60	0·64

[a] Cox (1966)
[b] McPhie (1966)

The mode of biosynthesis of ribosomal RNA may be contrasted in animals and bacteria. The evidence suggests that ribosomal RNA in animal cells is derived from precursors about twice the size of the mature molecules (Perry, 1962; Scherrer *et al.*, 1963; Rake and Graham, 1964; Penman, 1966; Perry, 1967; Warner and Soeiro, 1967). While still in the nucleus the precursor appears to be split into one or two molecules of ribosomal RNA plus some low molecular weight "waste" products which are not transported to the cytoplasm (Jeanteur, *et al.*, 1968). The function of the latter, representing one to two-thirds of the mass of the precursor, is unknown. The synthesis of ribosomal RNA in bacteria seems to proceed in a more straightforward manner since no high molecular weight precursors have been found to date.

Several workers have demonstrated the multiplicity of genes within a single cell which code for ribosomal RNA (Spiegelman and Yanofsky, 1965; Attardi *et al.*, 1965a, b; Birnstiel *et al.*, 1966; Dubnau *et al.*, 1965a, b; McConkey and Hopkins, 1964; Oishi and Sueoka, 1965; Ritossa *et al.*, 1966; Ritossa and Spiegelman, 1967; Takahashi *et al.*, 1967; Wallace and Birnstiel, 1966; Yanofsky and Spiegelman, 1963a, b). *E. coli* with a DNA content of 0·4 μμg contains approximately 10 genes for each of the two components of ribosomal RNA. In reticulocytes with a DNA content of 6 μμg DNA there are estimated 1 600–2 400 genes for each of the two components (Wallace and Birnstiel, 1966). Whether or not these genes have evolved separately, resulting in a heterogeneous population of ribosomes within a single cell, has been the subject of intense speculation. As yet no direct evidence has been

presented concerning the existence of sequence heterogeneity within ribosomal RNA from any organism. Some of the results described below are relevant to this question.

If ribosomal RNA has been modified in the course of evolution by mutation and selection of ribosomal genes, this should of course be directly indicated by changes in the nucleotide sequence. Hence it is of interest to study the nature and position of such changes and their effects on the structure and function of the ribosome.

The sequences of five transfer RNAs (Holley *et al.*, 1965; Zachau *et al.*, 1966, 1967; Madison *et al.*, 1967; and Raj Bhandary *et al.*, 1967) and 3 5S ribosomal RNAs (Brownlee *et al.*, 1967; Forget and Weissman, 1967) have been determined, but these are much smaller molecules (77–86 and 120 nucleotides, respectively) than the components of ribosomal RNA (approximately 3 000–5 000 and 1 500–2 000 nucleotides, respectively, for the larger and smaller components). Presumably it will be a long time before the sequences of these are known. One must therefore seek a simpler method for revealing sequence differences between high molecular weight RNA species. For this purpose we have developed what I shall term the "fingerprinting" method. This method appears to be particularly sensitive to the fundamental conformational nature of the RNA, which, within limits to be defined, overrides details of sequence alone.

"FINGERPRINTING" RNA

The basis of the method is the observation that the RNA chain is preferentially cleaved by certain nucleases in a relatively few and well-defined positions (McPhie *et al.*, 1966; Gould, 1966a, b). In the early stages of digestion the pattern of fragmentation is found to be quite characteristic of the RNA species. To analyse the resultant mixture of fragments a fractionation procedure of high resolution, sensitive essentially to molecular size, is required. Neither analytical centrifugation (Huppert and Pelmont, 1962) nor gel filtration (Zachau, 1965) offers anything like sufficient resolution and sensitivity. We have found that electrophoresis in polyacrylamide gels best meets the requirements.

As applied to proteins, gel electrophoresis fractionates both according to the size and charge of the molecules. With RNA, however, only the size is of relevance because all RNAs have the same ratio of charge to mass, there being one negatively charged phosphate group per residue. Therefore, it is possible to obtain an estimate of the size of RNA from its mobility on polyacrylamide gel electrophoresis once the relation of mobility to molecular weight or other suitable criterion of size is experimentally established (see below).

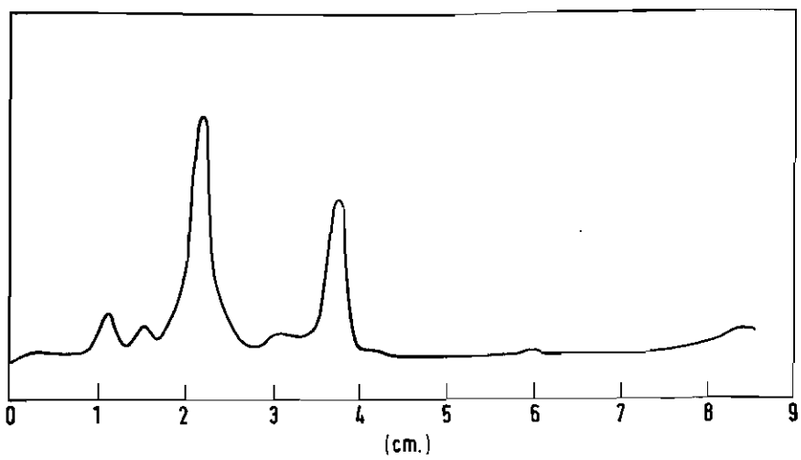


FIG. 1(*a*). Fractionation of rabbit reticulocyte ribosomal RNA on 2·2% polyacrylamide gels. Experimental details are described by Loening (1966). Comparable separations of ribosomal subunits have been reported by Hjerten *et al.* (1965).

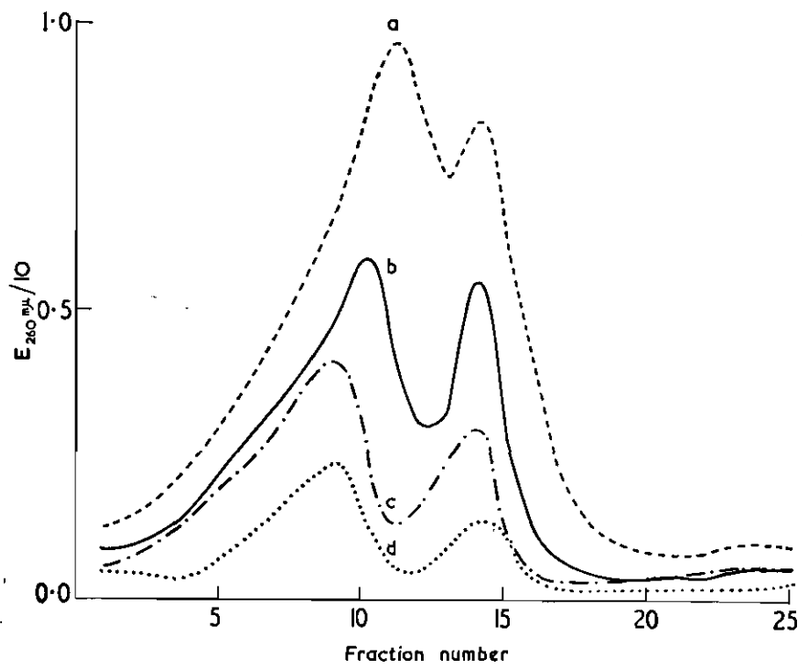


FIG. 1(*b*). Fractionation of rabbit reticulocyte ribosomal subunits by zonal centrifugation in sucrose gradients. The curves from bottom to top represent increasing loads of material applied to the gradient under otherwise constant conditions. Experimental details are given by Klucis and Gould (1966).

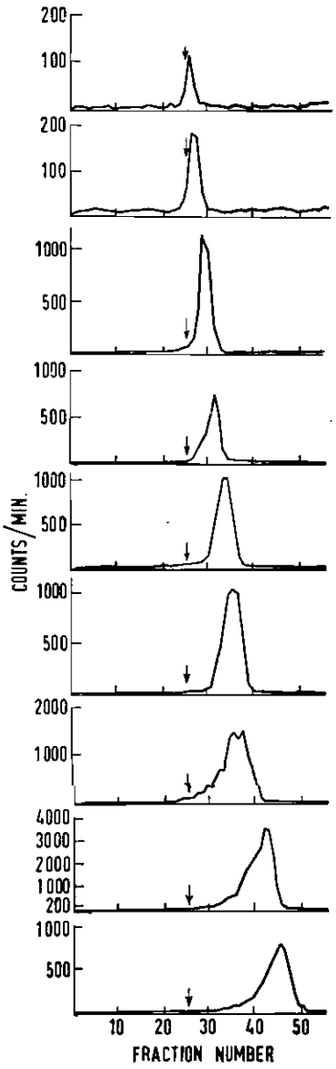


FIG. 2. Migration of oligonucleotides of defined length on 15% polyacrylamide gels. The oligonucleotides were obtained from a pancreatic ribonuclease digest of ^{32}P- and ^{3}H-labelled $\mu 2$ viral RNA and separated on a DEAE cellulose column containing 8M urea. Further experimental details are given by Gould *et al.* (in press).

We have found that gel electrophoresis is applicable over an enormous range of size extending from molecular weights of several millions down to mononucleotides. The concentration of acrylamide monomer used to make up the gels determines the effective pore size and hence the sieving properties of the gel. To achieve optimal resolution of particular RNA

species the appropriate gel concentration must be selected. For example, very dilute 2·5% polyacrylamide gels may be used to separate the components of ribosomal RNA (Fig. 1*a*; cf. *b* for comparable separation by zone centrifugation in sucrose gradients), while more concentrated 15% gels must be used to separate mono- and di-nucleotides (Fig. 2). To separate all the components of a complex

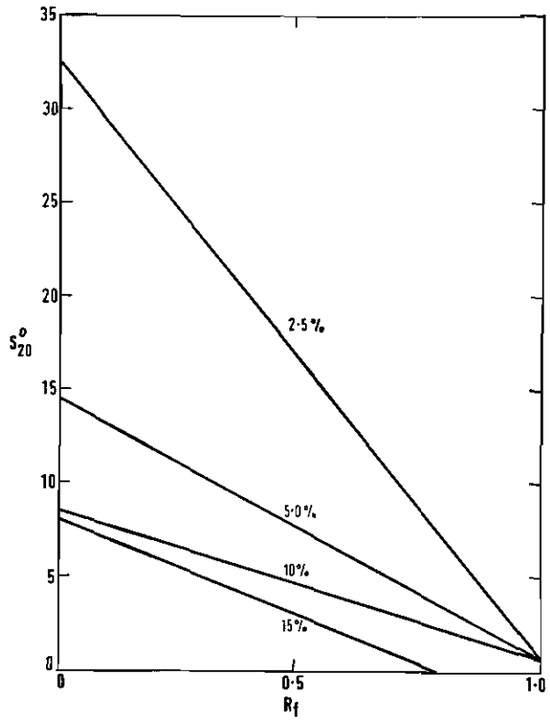


FIG. 3. The relation between electrophoretic mobility and sedimentation coefficient for polyacrylamide gels of various concentrations. R_f is the mobility of RNA relative to the anion (Cl^-) front in the discontinuous tris-barbiturate system recommended by Richards *et al.* (1965).

mixture several runs in gels of different concentrations may be required. Criteria for selecting the appropriate gel concentration are described here.

It has been shown experimentally that the relative mobility of RNA in polyacrylamide gel electrophoresis is linearly related to the sedimentation coefficient $S^\circ_{20,w}$ over a wide range of gel concentration. The

results of four independent calibrations of R_f in terms of $S^{\circ}_{20,w}$ for gels of various concentrations are shown in Fig. 3 (Richards *et al.*, 1965; McPhie *et al.*, 1966; Gould *et al.*, in press; Bishop and Spiegelman, 1967; Gould, unpublished). The molecular weight (M) of randomly coiled linear polymer chains is in turn related to $S^{\circ}_{20,w}$ by an equation of the type $S^{\circ}_{20,w} = kM^a$. The values of k and a for fragments of calf liver ribosomal RNA produced by thermal degradation are 2×10^{-2} and 0·5, respectively (Hall and Doty, 1958). Thus log M is likewise related to the electrophoretic mobility of RNA. It should be pointed out, however, that the exact values of k and a will in general depend to some extent on the conformation of the RNA so that the molecular weights determined in this manner must be interpreted with caution.

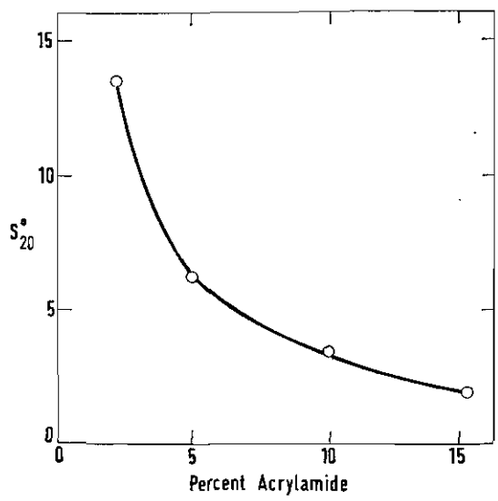


FIG. 4. Optimal acrylamide concentration for resolving RNA species. Optimal resolution of two RNA species occurs when their mean $R_f = 0{\cdot}6$ (Richards *et al.*, 1965). Sedimentation coefficients corresponding to $R_f = 0{\cdot}6$ for gels of various concentration taken from Fig. 3 are plotted here against the gel concentration.

Optimal resolution of two RNA species seems to occur when their mean R_f is around 0·6 (Richards *et al.*, 1965). As a rough guide Fig. 4 (based on Fig. 3) shows the gel concentration which should be selected to separate any two given RNA species close in size. For most purposes, however, this choice is not critical as long as the concentration is sufficiently low to allow the RNA to enter the gel and sufficiently high to give an appreciable pore-size effect (Fig. 3).

At least two factors account for the higher resolution which this

method affords, compared with other methods. First, despite the application of relatively large sample volumes, the RNA enters the gel in a very narrow zone by virtue of a sharpening effect operative in the discontinuous buffer system which we employ. (Ornstein, 1964; Williams and Reisfeld, 1964.) A second advantage is the lack of appreciable diffusion of the RNA as shown by the following experiment. After electrophoresis of three identical samples, one gel was stained immediately, the second after 3 h, and the third after 16 h (Fig. 5). Diffusion is seen to be slight for all but the fastest migrating species.

Zones are detected by staining and removing excess dye (Richards *et al.*, 1965), by direct ultra-violet scanning (Loening, 1966; Tsanev, 1965), or, if the samples are radioactive, by autoradiography (Fig. 6; Fairbanks *et al.*, 1965) or by "fractionating" the gel and measuring the radioactivity of fractions (Maizel, 1966). These methods are very sensitive: a few μg in a single zone can easily be detected by staining with acridine orange or pyronine Y. The method offers preparative possibilities in principle (see Hjerten *et al.*, 1965), but up to the present time no systematic attempt has been made to explore the limits of resolution employing fractionation of RNA on a significantly larger scale. Preliminary experiments with the apparatus described by Gordon and Louis (1967) have given promising results (Gould *et al.*, in press).

For the isolation of RNA from zones we have resorted to prestaining the RNA and, after electrophoresis, excising the prestained zones, grinding the slices in buffer to extract the RNA and removing the polyacrylamide by centrifugation (Gould, 1967). Although this method is time consuming and involves several manipulations, the extracted RNA was apparently undegraded.

TAXONOMIC RELATIONSHIPS

The first use of this method in "fingerprinting" revealed the existence of striking differences in the nuclease digestion patterns of ribosomal RNAs from various species (Gould *et al.*, 1966). In this study unfractionated ribosomal RNAs from *E. coli*, yeast, *Xenopus laevis* oöcytes, pigeon reticulocytes, mouse myeloma oöcytes tumour, rabbit liver and rabbit reticulocytes were compared (Fig. 7). All the samples were digested with T_1 ribonuclease under standard conditions and the digests were subjected to electrophoresis on 5% polyacrylamide gels. It will be recalled (Fig. 3) that 5% gels resolve RNAs in the range of 2·5–15S (corresponding molecular weights approximately $1{\cdot}5 \times 10^4$–$5{\cdot}5 \times 10^5$). Approximately half the mass of 30S RNA from rabbit reticulocytes is resolved into fragments in this molecular weight range under these conditions (Gould, 1966b). Smaller fragments including the remaiuder of 30S

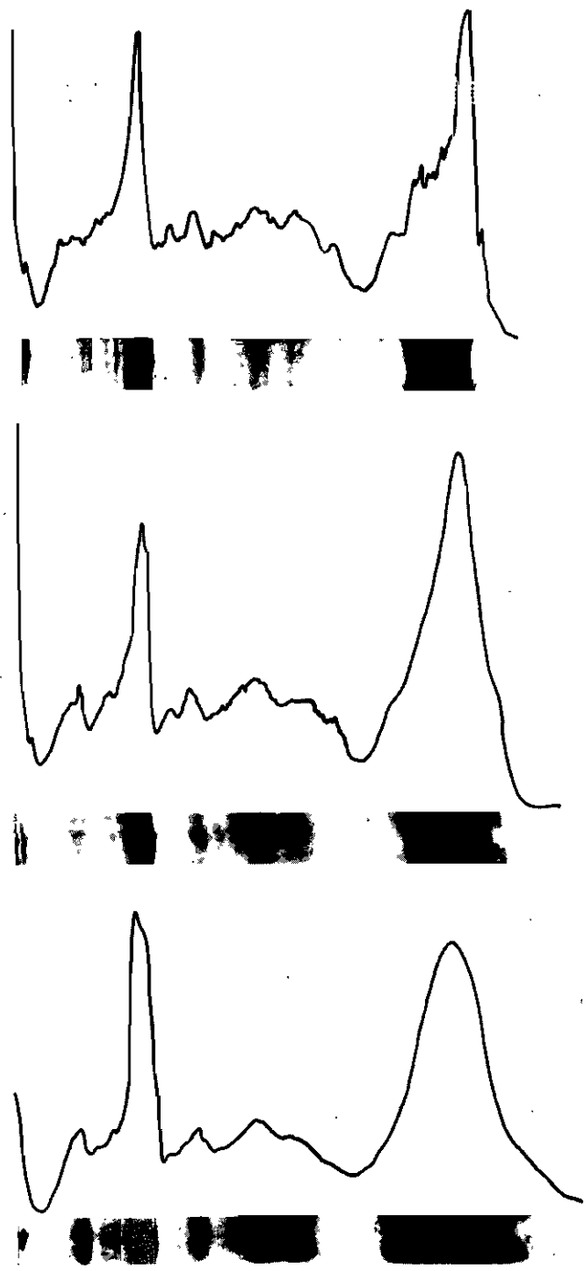
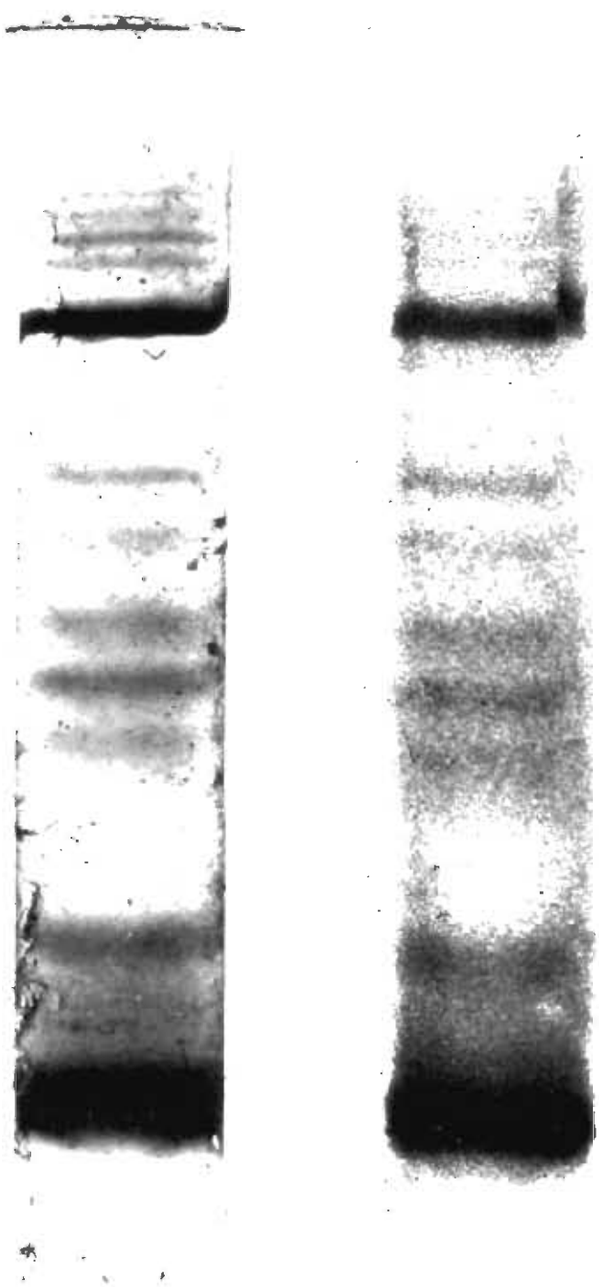


FIG. 5. Diffusion of RNA in polyacrylamide gels. Three samples of T_1 ribonuclease digest of rabbit reticulocyte ribosomal RNA were subjected to electrophoresis on 5% polyacrylamide gels in the tris-harbiturate system (Richards *et al.*, 1965). After electrophoresis for 40 min the first gel (top) was stained immediately, the second gel (middle) after 3 h and the third gel (bottom) after 16 h. Densitometer traces are shown above the photographs of the stained gels.



(a) (b)

FIG. 6. (legend facing)

RNA and fragments of 18S RNA run with the solvent front. They can be separated on 15% polyacrylamide gels (see below). The experiment clearly indicated that RNA from micro-organisms is more sensitive to nucleases than is that of higher organisms in that the fragments are consistently smaller. Differences in detail could also be seen on comparing the five ribosomal RNAs from the animal species. Ribosomal RNAs isolated from the two rabbit tissues, liver and reticulocytes, gave indistinguishable patterns, as might be expected from cells having the

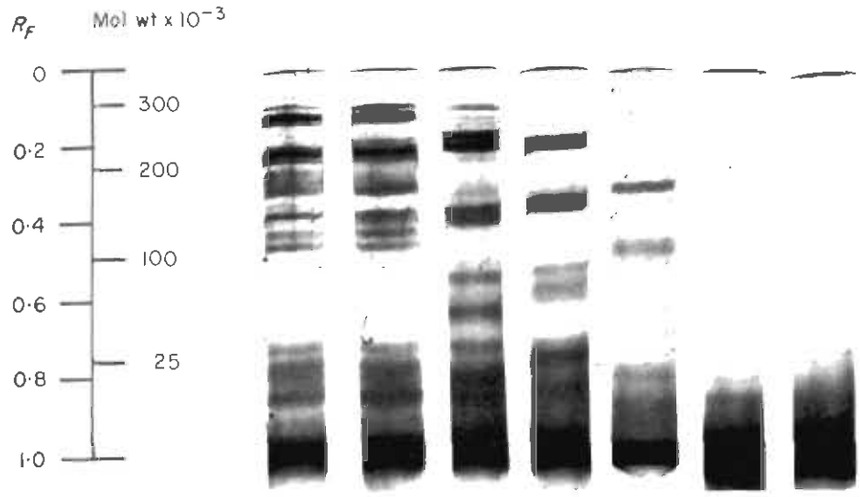


FIG. 7. Polyacrylamide gel electrophoretic patterns of T_1 ribonuclease digests of ribosomal RNA from various species. RNA was digested under standard conditions and electrophoresis carried out as described before (Gould *et al.*, 1966). From left to right: rabbit reticulocyte, rabbit liver, mouse ascites, pigeon reticulocyte, *Xenopus laevis* egg, *Saccharomyces fragilis* and *E. coli* ribosomal RNA.

FIG. 6. (p. 142) Radioautography of ^{14}C-labelled RNA after disc electrophoresis on polyacrylamide gel. Uniformly ^{14}C-labelled rabbit reticulocyte ribosomal RNA was digested with T_1 ribonuclease and electrophoresis was carried out as described before (Gould, 1966*a*). After staining and photography, the gel was sliced and radioautograms were prepared (Fairbanks *et al.*, 1965). The specific activity of the RNA was 8×10^4 CPM/mg and 250 μg was used for electrophoresis. X-ray film was exposed for 6 weeks to obtain the observed density of blackening in the radiogram (b); the pattern of staining with pyronine Y (a) is shown for comparison.

same genetic constitution. Reticulocyte ribosomal RNAs isolated from the phylogenetically distant species, pigeon and rabbit, on the other hand could clearly be distinguished on the basis of their "fingerprints".

We have recently been able to strengthen and extend these observations. Through the use of 15% polyacrylamide gels it has been possible to analyse the products of fragmentation of the larger component of ribosomal RNA in their entirety, to examine the "fingerprints" given by the smaller RNA component from various organisms, and to compare ribosomal RNAs from microbial species. The following preliminary results may be reported (Gould, Pinder, and Smith, unpublished).

Digestion of 30S RNA from rabbit reticulocytes and from liver ribosomes isolated from several mammalian species gave in addition to specific fragments of high molecular weight a series of relatively low molecular weight fragments ($< 10\,000$) resolvable in 15% polyacrylamide gels. The digestion patterns of rabbit reticulocyte and rabbit liver 30S ribosomal RNA were similar, as in the case of the high molecular weight fragments (Gould *et al.*, 1966). The patterns displayed by the three species, rat, mouse, and rabbit liver 30S RNA showed small but reproducible differences, particularly in the high molecular weight range.

Whereas the fragments of 19S ribosomal RNA from rabbit reticulocytes were for the most part too small to be resolved in 5% polyacrylamide gels, a series of well-resolved zones appeared on electrophoresis of a T_1 digest in 15% gels. In contrast to the case of 30S RNA, the patterns seen on comparing 19S RNA from rat, mouse and rabbit liver ribosomes were indistinguishable.

Ribosomal RNA was isolated from two species of yeast ribosomes, *Saccharomyces fragilis* and *Saccharomyces cerevisiae*. The RNAs were digested with three enzymes of different specificities, pancreatic ribonuclease (which attacks only pyrimidine residues), T_1 ribonuclease (only guanosine) and an enzyme purified from *Saccharomyces cerevisiae* (attacking preferentially the adenosine residues). The digestions were performed over a wide range of enzyme concentrations using standard incubation conditions (10 min at 25°C). Each of the three enzymes yielded a characteristic pattern with both 17S and 28S components of yeast ribosomal RNA. However, the RNAs from the two strains were indistinguishable under any of the conditions we employed.

Similar results were obtained with the two closely related bacterial species *B. subtilis* and *B. niger*. Although T_1 ribonuclease digestion patterns of 16S and 23S components differed appreciably, RNAs from the two species could not be distinguished.

HYPOTHETICAL NATURE OF RIBOSOMAL RNA MUTANTS AND METHODS OF STUDY

The results with *B. subtilis* and *B. niger* were of particular interest in view of the hybridization studies carried out by Dubnau *et al.* (1965a, b) in which the "homology" between these two species was found to be only 50% (rows 1–3 and columns 4 and 5, Table 5). It may be supposed

TABLE 5. Interstrain and interspecific hybridization with *Bacillus subtilis* RNA fractions.

DNA	% G + C	4S	16S	23S	Pulse	*In vitro*
B. subtilis SB19 $S^rE^rM^r$	43	100	100	100	100	100
					100	100
B. subtilis 6633 $E^rM^rS^r$	43	89	121	107	19	26
					22	27
B. niger E^rS^r	43	40	52	56	14	—
					16	—
B. pumilus M^rS^r	40–41	52	75	71	4	11
					11	11
B. licheniformis M^rS^r	45–46	87	118	113	23	—
					23	—
B. megaterium E^rS^r	39–40	65	121	118	0	10
					1	20
B. polymyxa S^r	47–48	26	43	36	1	—
					1	—
B. macerans S^r	52–53	47	78	78	7	3
					—	—
E. coli W	50	16	19	20	—	2
					0	*0*
Phage T_4	34	0	0	0	—	—

Experimental details described by Dubnau *et al.*, 1965b.

that this figure represents an upper limit, indeed probably a very considerable over-estimate, in sequence differences between the two species for the following reason: if two mutations in the RNA corresponding to ribosomal RNA are sufficiently close together (separated by less than about six nucleotides) then owing to the instability of short double helices it will not be possible for the RNA–DNA hybrid to form between these points. Thus eight rather than two phosphodiester linkages will be digested in the ribonuclease step. The hybridization technique is therefore capable of amplifying sequence differences between two ribosomal RNAs.

Any restrictions on the variation of base sequences in ribosomal RNA would tend to increase the probability of finding any two mutations close together and hence the discrepancy between the genetic homology

as measured by the hybridization technique and the actual number of base sequence changes.

It is easy to see how such restrictions might arise. The function of the ribosome in protein synthesis is undoubtedly dependent on the conformation of the ribosomal RNA constituents. In practice this entails the relative orientation of the numerous short helical regions of ribosomal RNA (Cox, Gould and Barkai, unpublished). A single mutation in any of the paired residues (which constitute 75% of the total) would result in mismatching according to the Watson-Crick base pairing scheme and could prevent helix formation. If the resultant ribosome were inactive or less efficient in protein synthesis, natural selection would tend to eliminate organisms bearing mutations at points corresponding to such paired residues in ribosomal RNA.

Further restrictions on the variability of base sequence might be envisaged if the precise sequence in either paired or unpaired regions is important for the interaction with ribosomal proteins, methylating enzymes or nucleases engaged in the biosynthesis of ribosomes or the supernatant proteins, nucleic acids (5S, ribosomal RN4, messenger and transfer RNAs) and coenzymes which participate in protein synthesis. Because of these likely restrictions one might predict that variations in sequence will be found to occur in relatively restricted portions of the sequence mainly situated in the unpaired or amorphous regions where the precise base composition may be unimportant.

The multiplicity of ribosomal RNA cistrons, if these are independent, may promote evolution by allowing mutations to accumulate in nonfunctional ribosomal RNA molecules. These molecules may be used with selective advantage subsequently in a different environment or when a number of mutations have occurred in other gene products which interact with ribosomes in protein synthesis.

Unlike the hybridization technique, "fingerprinting" appears to be sensitive primarily to the conformation of the RNA. It is likely that the conformation is determined in large degree by the nucleotide sequence, but that it can accommodate appreciable sequence variations, as in the case of proteins, particularly in unstructured regions. Since also the digestion would tend to occur preferentially in these latter regions it is not surprising that similar "fingerprints" are obtained on comparing closely related species. We conclude that mutations do not result in drastic changes in the overall structure of ribosomal RNA, but such changes may evolve gradually.

HOMOGENEITY OF RIBOSOMES IN A SINGLE ORGANISM

"Fingerprints" are capable in principle of yielding information about the homogeneity of ribosomes within a single organism. There are

approximately 2 000 cistrons for each component of ribosomal RNA in reticulocytes. The number of different ribosomal RNA sequences could be correspondingly large. Under certain circumstances one might therefore anticipate a very complex "fingerprint" with fragments in very

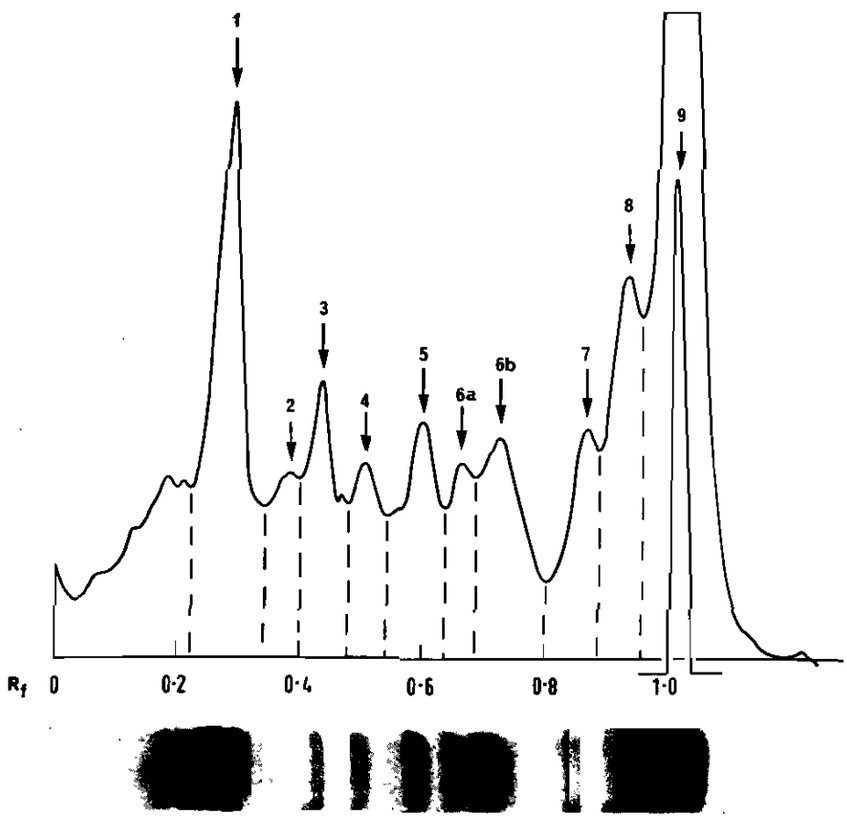


FIG. 8. Microdensitometer tracing of polyacrylamide gel shown in Fig. 9 (0). The R_f values reported in Table 2 are calculated from peak distance from the origin relative to that of the unresolved material (R_f = 1). The areas indicated beneath the peaks relative to the total area under the curve give an approximate estimate of the weight fraction of the fragment in the digest.

low molar yields compared to the intact RNA. That sequence changes can in fact be revealed by "fingerprints" is evidenced by the comparison of ribosomal RNAs from various species (Fig. 7).

In order to test whether ribosomal RNA from a single organism is

homogeneous, we have examined the molar yields of nuclease fragments. Figure 8 shows the way in which this information was obtained from a microdensitometer tracing of a gel containing a T_1 ribonuclease digest of rabbit reticulocyte ribosomal RNA stained with acridine orange. Since the fragments appear in molar yields approaching unity the RNA would appear by this criterion to be homogeneous (Gould, 1967). We have now obtained similar results on analyzing low molecular weight fragments of 30S and 19S ribosomal RNA from reticulocytes and from the three microbial species examined on 15% gels.

These observations do not rigorously prove that the ribosomes within a given organism are homogeneous for, as we have seen in the case of *B. subtilis* and *B. niger*, sequence differences can be masked by the overall conformational similarity of different RNAs. Moreover, the technique is not sufficiently sensitive to detect a small minority of mutant RNAs. It is of interest however that Birnstiel (personal communication) has obtained independent evidence based on the kinetics of formation of ribosome RNA–DNA hybrids that ribosomal RNA is homogeneous.

PROPERTIES OF RIBOSOMAL RNA FRAGMENTS

In order to pursue the structural differences between ribosomes further, we are now attempting to discover something about the nature of the fragments which we obtain. With this object we have isolated fragments of 30S rabbit reticulocyte ribosomal RNA (Fig. 9) and examined their thermal melting profiles (Table 6). The melting curves provide

TABLE 6. Properties of isolated ribosomal RNA fragments from rabbit reticulocytes.

Fragment	Yield (μg)	S° 20	$\frac{A_{280}}{A_{260}}$	$\frac{A_{235}}{A_{260}}$	ΔA^{rel}_{260}	ΔA^{rel}_{280}	$\frac{\Delta A^{rel}_{260}}{\Delta A^{rel}_{280}}$	$\frac{GC}{GC + AU}$
1	104	9·7	0·587	0·605	0·282	0·314	0·314	79
2	74	—	0·500	0·787	0·252	0·277	0·804	82
3	70	—	0·513	0·775	0·218	0·261	0·835	81
4	67	—	0·575	0·710	0·173	0·215	0·805	82
5	104	—	0·540	0·720	0·180	0·227	0·893	79
6	136	—	0·570	0·642	0·164	0·188	0·870	80
7	134	—	0·575	0·710	0·168	0·187	0·900	79
8	243	4·0	0·523	0·675	0·209	0·207	1·01	73
9	1 750	2·0	0·482	0·605	0·145	0·112	1·29	63
30S RNA	·	30·0	0·515	0·553	0·290	0·249	1·16	68

information on the stability and the composition of the helical regions of RNA. The GC content of the helical regions is obtained approximately from the equation

$$f_{GC} = 1/[0{\cdot}81(\Delta A_{260}/\Delta A_{280}) - 1{\cdot}46]$$

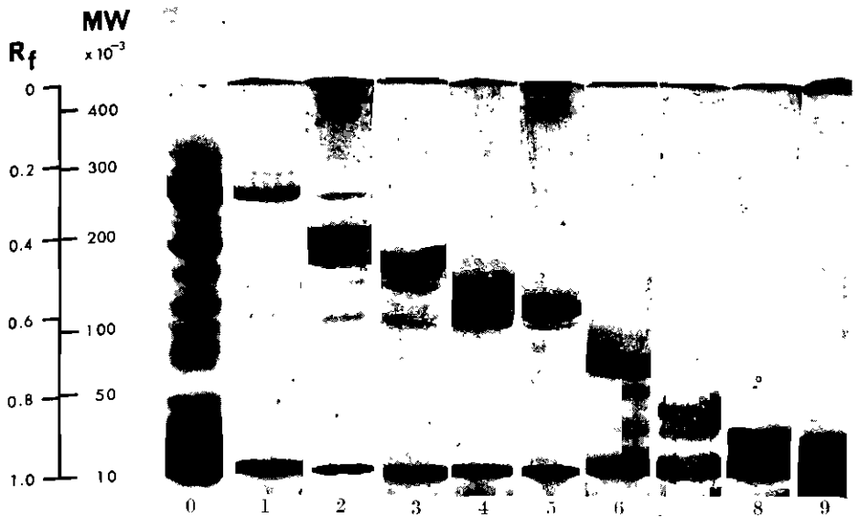


FIG. 9. Isolation of ribosomal RNA fragments. Ribosomal RNA was digested with T_1 ribonuclease. Pyronine *y* was added to stain the RNA and the digest was then subjected to electrophoresis in 5% polyacrylamide gel. The electrophoretic pattern obtained is shown at the far left together with the scale relating the mobility to molecular weight (McPhie *et al.*, 1966). RNA was extracted from nine separate zones and re-run on polyacrylamide gels as shown to the right of the original pattern. Further experimental details are reported in Gould (1967).

TABLE 7. Molar yields of ribosomal RNA fractions from rabbit reticulocytes.

Fragment (F)	Weight fraction	R_f	Molecular weight (MW)	$\frac{MW}{1{\cdot}6 \times 10^6}$	$\frac{\text{Moles F}}{\text{Moles 30S}}$
1	0·161	0·294	254 000	0·159	1·00
2	0·043	0·380	206 000	0·129	0·33
3	0·065	0·435	178 000	0·111	0·58
4	0·059	0·500	148 000	0·093	0·68
5	0·076	0·594	112 000	0·070	1·08
6a	0·042	0·650	92 000	0·058	0·72
6b	0·089	0·710	72 000	0·045	1·98
7	0·069	0·850	36 000	0·022	3·10
8	0·099	0·920	22 000	0·014	7·00
9	0·575	1·000	10 000	—	—
30S RNA	—	—	1 600 000	1 000	—

(Cox, 1966). The high melting temperature (T_M) (Fig. 10) and the GC content (Table 7) of the fragments of 30S ribosomal RNA isolated from 5% polyacrylamide gels (Fig. 9) is reminiscent of the upper melting phase of the intact RNA (Fig. 10:0). Since also the proportion (about 40%) of the total RNA is consistent, it is reasonable to conclude that we have thus isolated these regions intact.

The thermostable GC-rich sequences of ribosomal RNA from higher organisms therefore seem to be concentrated in confined regions of the molecule. It will be recalled that RNA from higher organisms is 20–40% bigger than that of bacteria (Table 1). We suppose therefore that insertions were made at specific points (or additions at the end) of the ribosomal genes in the course of evolution. It would be of interest to determine the relative position of these new sequences both in terms of the primary sequence of the RNA and the topology of the ribosome.

COMMENTS ON THE EVOLUTION OF RIBOSOMAL PROTEINS

We may briefly consider what may be deduced about the evolution of the ribosomal proteins. Multiplying the molecular weight of the ribosome by the percentage of protein in the ribosome (Table 1) we obtain the total contribution of the ribosomal proteins to the molecular weight of the ribosome. The values calculated in this way are $1{\cdot}2 \times 10^6$ for *E. coli*, $2{\cdot}5 \times 10^6$ for yeast and $2{\cdot}7 \times 10^6$ for rat liver ribosomal proteins. Closely similar values are obtained by subtracting the molecular weight of the RNA components (Table 2) from that of the corresponding ribosomes (Table 1). The average molecular weights of the ribosomal proteins of *E. coli* (M_w = 23 000: Möller and Chrambach, 1967) and of rat liver (M_w = 26 000: Hamilton and Ruth, 1967) are essentially identical. Therefore the number of proteins in *E. coli* is 50–65 and 100 in rat liver. Thus about 35–50 new proteins have been added to ribosomes in the course of evolution. Thus both the length of RNA and the number of proteins have increased on the evolution of higher organisms from bacteria. It is tempting to speculate that the "young" proteins and RNA are associated together and the two evolved in tandem.

FUTURE PROSPECTS

The more difficult problem for the future is to elucidate the mechanism of protein synthesis in all essential detail, and to interpret in mechanistic terms the significance of the evolutionary changes. Much is already known about the mechanism of protein synthesis (Schweet and Heintz, 1966; Attardi, 1967). However, though the structural features which differentiate ribosomes appear to be correlated with

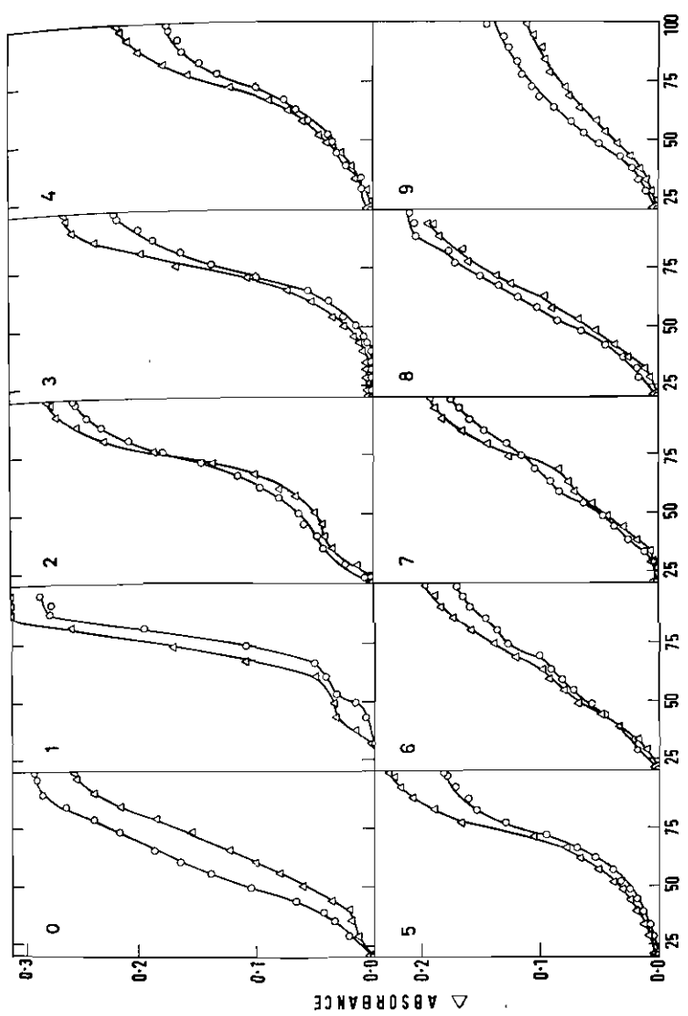


FIG. 10. Melting profiles of the fragments shown in Fig. 9 were dissolved in 0·01 M phosphate buffer, pH 7·6, and the absorbance changes on heating were measured as described by Gould (1967). The melting profile of the 30S component of rabbit reticulocyte ribosomes is shown in the upper left panel for comparison.

modifications in the exact mode of function (Takeda and Lipmann, 1966; Küntzell and Noll, 1967; Parisi *et al.*, 1967) the uature and selective advantage of the functional differences remain to be discovered.

From the taxonomic viewpoint the study of RNA seems to have a number of advantages, in particular, that unlike most proteins, and certainly most experimentally accessible proteins it is a very "old" molecule which is common in its occurrence and function to all forms of life. The work which I have presented should be viewed as only the beginning of the arduous road to a fuller understanding of the structure and function relationship in ribosomes, and the differences which evolution has generated.

ACKNOWLEDGMENTS

Some of the unpublished work which I have discussed was performed in collaboration with Miss Jennifer Pinder. The work on *B. subtilis* and *B. niger* RNA is part of a study undertaken with Dr. Julius Marmur and Dr. Isaar Smith. I am grateful to Dr. M. Spencer who provided the samples which were used in the work on yeast ribosomal RNA. My sincere thanks are due to Professor Sir John Randall and Professor M. F. H. Wilkins for their encouragement.

Figures 7, 8, 9 and 10, and Tables 6 and 7 are reproduced by kind permission of the editor of the *Journal of Molecular Biology*. Table 5 is reproduced by kind permission of the National Academy of Sciences, U.S.A.

REFERENCES

Attardi, G. (1967). The mechanism of protein synthesis. *Ann. Rev. Microbiol.* 383–416.

Attardi, G., Huang, P. and Kabat, S. (1965a). Recognition of ribosomal RNA sites in DNA. I Analysis of the *E. coli* system. *Proc. natn. Acad. Sci. U.S.A.* **53**, 1490–98.

Attardi, G., Huang, P. and Kabat, S. (1965b). Recognition of ribosomal RNA sites in DNA, II. The HeLa cell system. *Proc. natn. Acad. Sci., U.S.A.* **54**, 185–92.

Birnstiel, M. L., Wallace, H., Sirlin, J. L. and Fischberg, M. (1966). Localisation of the ribosomal DNA complements in the nucleolar organizer region of *Xenopus laevis*. *Natn. Cancer Inst. Monogr.* **23**, 431–47.

Bishop, D. H. L. and Spiegelman, S. (1967). An analysis by gel electrophoresis of Qβ-RNA complexes formed during the latent period of an *in vitro* synthesis. *Proc. natn. Acad. Sci., U.S.A.* **57**, 1474–81.

Brownlee, G. C., Sanger, F. and Barrell, B. S. (1967). Molecular structure—Nucleotide sequence of 5S ribosomal RNA from *Escherischia coli*. *Nature, Lond.* **215**, 735–36.

Chao, F. C. and Schachman, H. (1956). The isolation and characterisation of a macromolecular ribonucleoprotein from yeast. *Archs. biochem. Biophys.* **61**, 220–30.

Cox, R. A. (1966). Spectrophotometric study of the conformation of 19S and 30S ribosomal ribonucleic acid from rabbit reticulocytes. *Biochem. J.* **98**, 841–57.

Dintzis, H. M., Boorsook, H. and Vinograd, J. (1958). Microsomal structure and haemoglobin synthesis in the rabbit reticulocyte. *In* Roberts (ed.) "Microsomal Particles and Protein Synthesis", pp. 95–99. Pergamon, Oxford.

Dubnau, D., Smith, J. and Marmur, J. (1965a). Gene conservation in Bacillus species II. The location of genes concerned with the synthesis of ribosomal components and soluble RNA. *Proc. natn. Acad. Sci., U.S.A.* **54**, 724–30.

Dubnau, D., Smith, J., Morrell, P. and Marmur, J. (1965b). Gene conservation in Bacillus species I. Conserved genetic and nucleic acid base sequence homologies. *Proc. natn. Acad. Sci., U.S.A.* **54**, 491–98.

Fairbanks, G., Levinthal, C. and Reeder, R. H. (1965). Analysis of C^{14}-labelled proteins by electrophoresis. *Biochem. biophys. Res. Commun.* **20**, 393–99.

Forget, B. G. and Weissman, S. M. (1967). Nucleotide sequence of K B cell 5S RNA. *Science, N.Y.* **158**, 1695–99.

Gierer, A. (1958). Vergleichende Untersuchungen an hochmolekularer Ribosenucleinsäure. *Naturforsch.* **13B**, 788–92.

Gordon, A. H. and Louis L. N. (1967). Preparative acrylamide electrophoresis. A single gel system. *Analyt. Biochem.* **21**, 190–200.

Gould, H. J. (1966a). The specific cleavage of ribonucleic acid from reticulocyte ribosomal subunits. *Biochemistry, N.Y.* **5**, 1103–08.

Gould, H. J. (1966b). A comparison of specific fragments produced by partial digestion of reticulocyte ribosomal RNA with pancreatic and T_1 ribonuclease. *Biochim. biophys. Acta* **123**, 441–44.

Gould, H. J., Bonanou, S. and Kanagalingham, K. (1966). Structural characterisation of ribosomal ribonucleic acids from various species by a new "Fingerprinting" technique. *J. molec. Biol.* **22**, 397–99.

Gould, H. J. (1967). The nature of high molecular weight fragments of ribosomal RNA. *J. molec. Biol.* **29**, 307–13.

Gould, H. J., Pinder, J., Matthews, H. and Gordon, A. H. (in press). The fractionation of low molecular weight fragments of ribosomal and viral RNA by polyacrylamide gel electrophoresis. *Analyt. Biochem.*

Hall, B. D. and Doty, P. (1958). The configurational properties of ribonucleic acid isolated from microsomal particles of calf liver. *In* Roberts (ed.) "Microsomal Particles and Protein Synthesis", pp. 27–35. Pergamon, Oxford.

Hamilton, M. G. and Ruth, M. E. (1967). Characterization of some of the proteins of the large subunits of rat liver ribosomes. *Biochem. J.* **6**, 2585–91.

Hastings, J. R. B. and Kirby, K. S. (1966). The nucleic acids of *Drosophilia melanogaster*. *Biochem. J.* **100**, 532–39.

Hjerten, S., Jerstadt, S. and Tiselius (1965). Electrophoretic "particle sieving" in polyacrylamide gels as applied to ribosomes". *Analyt. Biochem.* **11**, 219–23.

Holley, R. M., Merrill, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. and Zamir, A. (1965). Structure of a ribonucleic acid. *Science, N.Y.* **147**, 1462–65.

Huppert, J. and Pelmont, J. (1962). Ultracentrifuge studies of RNA degradation. *Archs. biochem. Biophys.* **98**, 214–23.

Ingram, J. M. (1958). Abnormal haemoglobins. I. The comparison of normal sickle-cell haemoglobins by "fingerprinting". *Biochim. biophys. Acta* **28**, 539–45.

Jeanteur, Ph., Amaldi, F. and Attardi, G. (1968). Partial sequence analysis of ribosomal RNA from HeLa cells. II Evidence for non-ribosomal-type sequences in 45S and 32S ribosomal RNA precursors. *J. Mol. Biol.* **33**, 757–70.

Klucis, E. and Gould, H. J. (1966). Zonal ultracentrifugation for the separation of ribosomal subunits. *Science, N.Y.* **152**, 378.

Küntzell, H. and Noll, H. (1967). Mitochondrial and cytoplasmic polysomes from *Neurospora crassa*. *Nature, Lond.* **215**, 1340–45.

Kurland, C. G. (1960). Molecular characterization of ribonucleic acid from *E. coli* ribosomes. *J. molec. Biol.* **2**, 83–91.

Loening, U. (1966). The fractionation of high-molecular weight RNA by polyacrylamide gel electrophoresis. *Biochem. J.* **102**, 251–57.

Madison, J. T., Everett, G. A. and King, H. K. (1967). On the nucleotide sequence of yeast transfer RNA. *Cold Spring Harb. Symp. quant. Biol.* **31**, 409–16.

Maeda, A. (1961). Some properties of ribonucleic acid from yeast 80S particles: Effect of pH and magnesium ions. *J. Biochem., Tokyo* **50**, 377–85.

Maizel, J. (1966). Acrylamide gel electrophorograms by mechanical fractionation: Radioactive adenovirus protein. *Science, N.Y.* **151**, 988–90.

McConkey, E. H. and Hopkins, J. W. (1964). The relationship of the nucleolus to the synthesis of ribosomal RNA in HeLa cells. *Proc. natn. Acad. Sci. U.S.A.* **51**, 1197–1204.

McPhie, P. (1966). Studies on yeast ribosomes. Ph.D. Thesis, London University.

McPhie, P., Hounsell, J. and Gratzer, W. B. (1966). The specific cleavage of yeast ribosomal ribonucleic acid with nucleases. *Biochemistry, N.Y.* **5**, 988–93.

Möller, W. and Chrambach, A. (1967). Physical heterogeneity of ribosomal proteins from *E. coli*. *J. molec. Biol.* **23**, 377–90.

Oishi, M. and Sueoka, N. (1965). Location of genetic loci of ribosomal RNA on *Bacillus subtilis* chromosome. *Proc. natn. Acad. Sci. U.S.A.* **54**, 483–91.

Ornstein, L. (1964). Disc electrophoresis I. Background and theory. *Ann. N.Y. Acad. Sci.* **121**, 321–49.

Parisi, B., Milanesi, G., Van Etten, J. L., Perani, A. and Ciferri, G. (1967). Species specificity in protein synthesis. *J. molec. Biol.* **28**, 295–309.

Penman, S. (1966). RNA metabolism in the HeLa cell nucleus. *J. molec. Biol.* **17**, 117–30.

Perry, R. P. (1962). The cellular sites of synthesis of ribosomal and 4S RNA. *Proc. natn. Acad. Sci. U.S.A.* **48**, 2179–86.

Perry, R. P. (1967). The nucleolus and the synthesis of ribosomes. *In* Davidson, J. N. and Cohn, W. E. "Progress in Nucleic Acid Research and Molecular Biology" Vol. 6, 219–57. Academic Press, New York.

Petermann, N. M. (1964). "The Physical and Chemical Properties of Ribosomes." Elsevier, Amsterdam.

Raj Bhandary, A. L., Chang, S. H., Stuart, A., Faulkner, R. D., Hoskinson, R. M. and Khorana, H. G. (1967). Studies on polynucleotides, LXVIII. The primary structure of yeast phenylalamine transfer RNA. *Proc. natn. Acad. Sci., U.S.A.* **57**, 751–58.

Rake, A. V. and Graham, A. F. (1964). Kinetics of incorporation of uridine-C^{14} into L cell RNA. *Biophys. J.* **4**, 267–84.

Richards, E. G., Coll, J. A. and Gratzer, W. B. (1965). Disc electrophoresis of ribonucleic acid in polyacrylamide gels. *Analyt. Biochem.* **12**, 452–71.

Ritossa, F. M., Atwood, K. C., Lindsley, D. L. and Spiegelman, S. (1966). On the chromosomal distribution of DNA complementary to ribosomal and soluble RNA. *Natn. Cancer Inst. Monogr.* **23**, 449–72.

Ritossa, F. M. and Spiegelman, S. (1967). Localisation of DNA complementary to ribosomal RNA in the nucleolus organizer of *Drosophila melanogaster*. *Proc. natn. Acad. Sci. U.S.A.* **53**, 737–45.

Scherrer, K., Latham, H. and Darnell, J. E. (1963). Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. *Proc. natn. Acad. Sci. U.S.A.* **49**, 240–48.

Schweet, R. and Heintz, R. (1966). Protein synthesis. *Ann. Rev. Biochem.* **35**, 723–58.

Spahr, P. F. and Tissières, A. (1959). Nucleotide composition of ribonucleoprotein particles from *Escherichia coli*. *J. molec. Biol.* **1**, 237–39.

Spencer, M., Fuller, W., Wilkins, M. F. H. and Brown, G. (1962). Determination of the helical configuration of RNA molecules by X-ray diffraction study of crystalline amino-acid transfer ribonucleic acid. *Nature, Lond.* **194**, 1014–20.

Spencer, M. and Poole, F. (1965). On the origin of crystallizable RNA from yeast. *J. molec. Biol.* **11**, 314–26.

Spiegelman, S. and Yanofsky, S. A. (1965). The relation of ribosomal RNA to the genome. *In* Bryson, V. and Vogel, H. J. "Evolving Genes and Proteins" 537–79. Academic Press, New York.

Spirin, A. S. (1963). Some problems concerning the macromolecular structure of ribonucleic acids. *In* Davidson, J. N. and Cohn, W. E. "Progress in Nucleic Acid Research" Vol. 1, 301–45. Academic Press, New York.

Stutz, E. and Noll, H. (1967). Characterization of cytoplasmic and chloroplast polysomes in plants. Evidence for three classes of ribosomal RNA in nature. *Proc. natn. Acad. Sci., U.S.A.* **57**, 774–78.

Takahashi, H., Saito, H. and Ikeda, Y. (1967). Species specificity of the ribosomal RNA cistrons in bacteria. *Biochim. biophys. Acta,* **134**, 124–33.

Takeda, M. and Lipmann, F. (1966). Comparison of amino acids polymerization in *B. subtilis* and *E. coli* cell-free systems: Hybridization of their ribosomes. *Proc. natn. Acad. Sci. U.S.A.* **56**, 1875–82.

Tashiro, Y. and Yphantis, D. A. (1965). Molecular weights of hepatic ribosomes and their subunits. *J. molec. Biol.* **11**, 174–86.

Taylor, M. M., Glasgow, J. E. and Storck, R. (1967). Sedimentation coefficient of RNA from 70S and 80S ribosomes. *Proc. natn. Acad. Sci. U.S.A.* **57**, 165–69.

Taylor, M. M. and Storck, R. (1964). Uniqueness of bacterial ribosomes. *Proc. natn. Acad. Sci. U.S.A.* **52**, 958–65.

Tissières, A., Watson, J. D., Schlesinger, D. and Hollingsworth, B. R. (1959). Ribonucleoprotein particles from *Escherichia coli*. *J. molec. Biol.* **1**, 221–33.

Tsanev, R. (1965). Direct spectrophotometric analysis of RNA fractionation by agar-gel electrophoresis. *Biochim. biophys. Acta* **103**, 374–82.

Ts'o, P. O. P., Bonner, J. and Vinograd, J. (1958). Structure and properties of microsomal nucleoprotein particles from pea seedlings. *Biochim. biophys. Acta* **30**, 570–82.

Wallace, H. and Birnstiel, M. (1966). Ribosomal cistrons and the nucleolar organizer. *Biochim. biophys. Acta* **114**, 296–310.

Warner, J. R. and Sociro, R. (1967). Nascent ribosomes from He La cells, *Proc. Nat. Acad. Sci. U.S.* **58**, 1984.

Williams, D. E. and Reisfeld, R. A. (1964). Disc electrophoresis in polyacrylamide gels: Extension to new conditions of pH and buffer. *Ann. N.Y. Acad. Sci.* **121**, 381–90.

Yanofsky, S. A. and Spiegelman, S. (1963a). The identification of the ribosomal RNA cistrons by sequence complementarity. I. Specificity of complex formation. *Proc. natn. Acad. Sci. U.S.A.* **48**, 1069–78.

Yanofsky, S. A. and Spiegelman, S. (1963b). The identification of the ribosomal RNA cistrons by sequence complementarity II. Saturation of and competitive interaction at the RNA cistron. *Proc. natn. Acad. Sci. U.S.A.* **48**, 1466–72.

Zachau, H. G. (1965). Oligo- und Polynucleotidtrennungen IV. Chromatographie von löslicher Ribonucleinsäure und deren Spaltprodukten an Sephadex. *Biochim. biophys. Acta* **108**, 355–66.

Zachau, H. G., Dütting, D. and Feldman, H. (1966). Nucleotide sequences of serine specific transfer ribonucleic acid. International edition, *Angew. Chem.* **5**, 422.

Zachau, H. G., Dütting, D., Feldman, H., Melcher, F. and Karn, W. (1967). Serine specific transfer ribonucleic acids XIV. Comparison of nucleotide sequences and secondary structure models. *Cold Spring Harb. Symp. quant. Biol.* **31**, 417–24.

SECTION IV

OTHER BIOCHEMICAL APPROACHES

15. Bile Salt Differences in Relation to Taxonomy and Systematics

G. A. D. Haslewood

Biochemistry and Chemistry Department, Guy's Hospital Medical School, London, England

ABSTRACT

Comparisons of the bile salts of representatives of vertebrate groups whose taxonomic affinities are not fully understood may be useful. Examples are hagfish and lampreys; holocephalans and selachians; primitive osteichthyeans and teleosts; crossopterygians, dipnoans and amphibians; certain lizards and snakes. Bile salts studies suggest that Ranidae are heterogeneous and biochemically more advanced than Bufonidae. They support the idea that chelonians have a long history separate from that of other reptiles; they may also be valuable in snake classification, may indicate an earlier (more vegetarian) history for some groups of mammals now having "unique" bile acids and might be valuable in investigating the "radiation" of eutherians having glycine conjugates.

INTRODUCTION

The use of a biochemical character for taxonomic purposes implies, above all, that the mechanistic nature of the character be understood. It is very well to say that "everything is in the DNA", but between the DNA and its biochemical expression lie complex controls and regulators that must be elucidated before the full meaning to taxonomy of the differences can be appreciated. Fortunately in the case of biochemical characters there seems no technical reason why this end cannot ultimately be attained.

In chemical bile salt differences we are considering the significance of variations in physiologically active molecules, secreted for a digestive purpose and shaped by evolution in most large vertebrate groups to a common pattern. We are not, as yet, in a position to be able to discuss

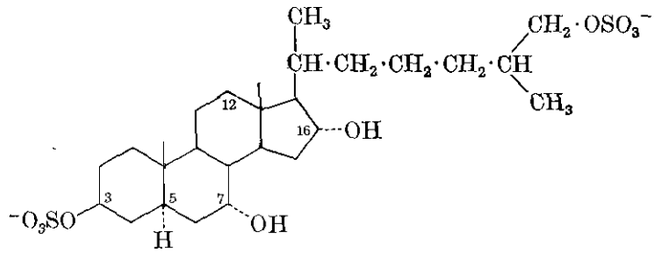
Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 159–172.

the mechanism of this evolution or even to examine the enzymes concerned in bile salt formation. This paper will therefore attempt to discuss chemical bile salt differences, as they exist, where possible taking into account the modifications in the primary bile salts (made in the liver from cholesterol) brought about by intestinal micro-organisms during the enterohepatic circulation.

Comparisons will be made in vertebrate groups to the taxonomic study of which bile salt differences may be relevant. Except where otherwise indicated, references to the chemistry and biochemistry of bile salts and to their occurrence in different species are given by Haslewood (1967a).

AGNATHA

According to Heintz (1962) "Kiaer and his supporters regard the recent cyclostomes as a monophyletic group and are of the opinion that the Petromyzontia and Myxinoidae diverged from one another comparatively recently". This seems to be a minority view amongst taxonomists and it receives no support from the bile salt chemistry. Myxinol disulphate, the chief bile salt in both *Myxine* and *Eptatretus* (*Bdellostoma*) is 3β,7α,16α,26(or 27)-tetrahydroxy-5α-cholestane 3,26 (or 27) disulphate (formula 1: Anderson *et al.*, 1967).

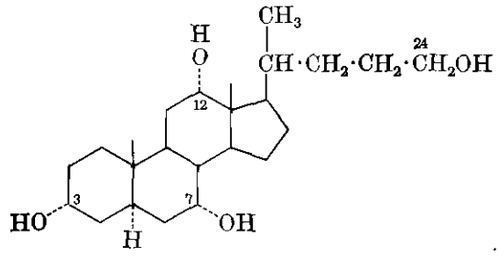


(1) Myxinol disulphate

Myxinol disulphate is the most primitive bile salt known, in the sense that it is biochemically nearest to cholesterol, having the 27 carbon atoms and 3β hydroxyl group of that substance and only four —OH groups in all. All other chief bile alcohol C_{27} sulphates have at least five hydroxyl groups (three in the 3,7,12 positions), only one sulphate ester group and, with the exception of latimerol (mentioned below), have the —OH group at C-3 in the α-configuration.

The chemistry of the bile salts of *Petromyzon marinus* has been elucidated to show that the chief component has a molecule with

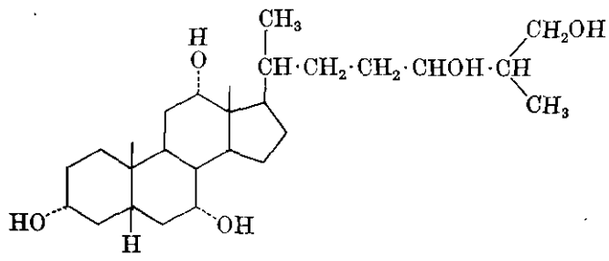
only one sulphate ester group, the 5α-configuration and three of its —OH groups in the 3α,7α,12α positions. Nuclear magnetic resonance (n.m.r.) and mass spectral examination of the bile alcohol ("petromyzonol") suggested 3α,7α,12α,24-tetrahydroxy-5α-cholane (2) as a possible structure. Chemical confirmation followed (Haslewood, and Tökés, 1968) and it seems clear that *Petromyzon* bile salts are entirely different from and also biochemically more advanced than those of the myxinids.



(2) Petromyzonol

CHONDRICHTHYES

Bile salt chemistry may be relevant to the relationships of the Holocephali to the Selachi. Patterson (1965) in discussing this question says that: "there is evidence of genetic affinity between the holocephalens and selachians, but no evidence of any direct relationship between the two groups". Between groups already recognized as related, similarities in bile salt chemistry might be taken to be indicative of the expression of homologous genes. In the present case, chimaerol (3α,7α,12α,24ξ,26(or 27)-pentahydroxy-5β-cholestane (Formula (3), now confirmed by n.m.r. examination by Dr. L. Tökés) has been found to be a principal bile alcohol of *Chimaera monstrosa* and also



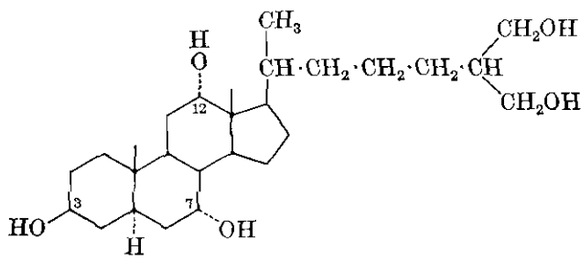
(3) Chimaerol

present in minor amounts in the bile of at least two selachians, the ray, *Dasyatis akajei*, and dogfish, *Mustelus manazo* (references in Haslewood, 1967b).

The chief bile alcohol in selachians seems to be scymnol, which differs from chimaerol only in having the terminal side-chain —$\overset{24}{C}HOH.\overset{25}{C}H(CH_2OH)_2$ instead of —$\overset{24}{C}HOH.\overset{25}{C}H(CH_3)CH_2OH$ as shown in Formula (3). If the steric configuration at C-24 is identical in chimaerol and scymnol (a point yet to be determined), it is hard to resist the conclusion that chimaerol is a biosynthetic precursor of scymnol and could give rise to it by the action of a single hydroxylating enzyme. Thus, in their bile salt chemistry (and probably biochemistry) the chimaeroids are primitive selachians. Their bile probably also contains a little scymnol.

PRIMITIVE OSTEICHTHYES

A chief bile salt in the coelacanth *Latimeria chalumnae* (now confirmed for a second specimen) is 3β,7α,12α,26,27-pentahydroxy-5α-cholestane (latimerol).

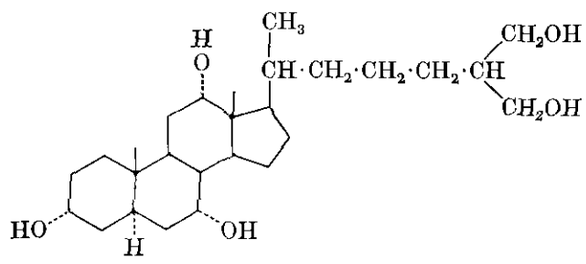


(1) Latimerol

Assuming that latimerol is biosynthesized from cholesterol and provided that the 3β-OH group is not a result of the intervention of intestinal micro-organisms (i.e. latimerol a secondary bile alcohol), this is a primitive substance, in accord with the position assigned to *Latimeria* by taxonomists. The 3α epimer of latimerol, 5α-cyprinol (5), also occurs in minor amounts in coelacanth bile and is a chief bile alcohol in dipnoans and certain freshwater teleosts.

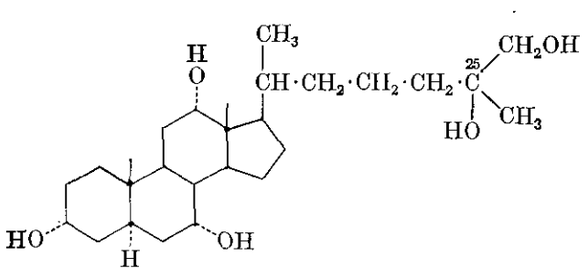
The latimerol/cyprinol hydroxylation patteru (3,7,12,26,27) in a 5α-cholestane seems to be characteristic of freshwater osteichthyeans; this fiuding provides further support for the views of paleontologists that the coelacanths, though now marine, were originally freshwater forms.

No latimerol has been detected in the bile of *Neoceratodus*, *Lepidosiren* or *Protopterus*; the lungfish are distinctly more advanced in their bile



(5) 5α-Cyprinol

salt chemistry than *Latimeria*. *Neoceratodus* bile salts are a mixture including the sulphate esters of 5α-cyprinol and 5α-bufol (3α,7α,12α, 25ξ,26(or 27)-pentahydroxy-5α-cholestane (6).



(6) 5α-Bufol

The bufol pattern of hydroxylation is found also in certain amphibia and is discussed later.

CHONDROSTEI

Sturgeon (three species of *Acipenser*) and paddlefish (*Polyodon*) bile salts agree in consisting mainly of taurocholic and tauroallocholic acids, as in most teleosts, and a small proportion of bile alcohol sulphates. The biochemical picture is of mainly advanced bile salts but with obvious primitive characters, in accord with the position assigned to chondrosteans by taxonomists.

TELEOSTEI

In all the instances cited above, there is no essential conflict between the taxonomic indications of bile salt chemistry and the views of

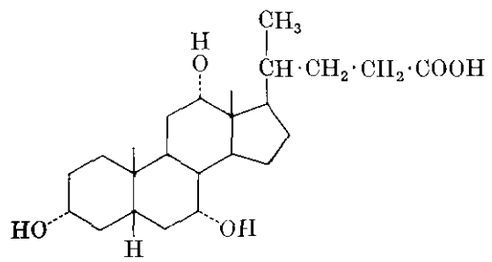
systematists enumerated on other grounds. However, among the teleosts bile salt indications do not always appear to agree with present-day taxonomic ideas.

For instance, in all the cyprinids examined (12 species of Cyprinidae and one each of Catostomidae and Cobitidae), the chief bile salt is the sulphate of 5α-cyprinol (5). This substance is undoubtedly a primitive bile alcohol.

If the teleostean biosynthetic stages between cholesterol (proved to be a source of the bile salts in Cyprinidae) and the bile salts are the same as those in mammals, the later steps so far elucidated probably are:

3α,7α,12α,26-tetrahydroxycholestane (A) → 3α,7α,12α-trihydroxycholestan-26-al (B) → 3α,7α,12α-trihydroxycholestan-26-oic acid (C) → 3α,7α,12α,24-tetrahydroxycholestan-26-oic acid (D) → 3α,7α,12α-trihydroxycholanoic acid (E) + propionyl CoA.

5α-reduction of the cholesterol double bond at a stage previous to substance (A) would lead to 5α-cyprinol via the 5α-epimer of (A), and 5β-reduction as above would give cholic acid (7) as the acid (E).



(7) Cholic acid

Thus, stripped of its chemical detail, the biosynthetic relationship between cholic acid and 5α-cyprinol may be as shown in Fig. 1.

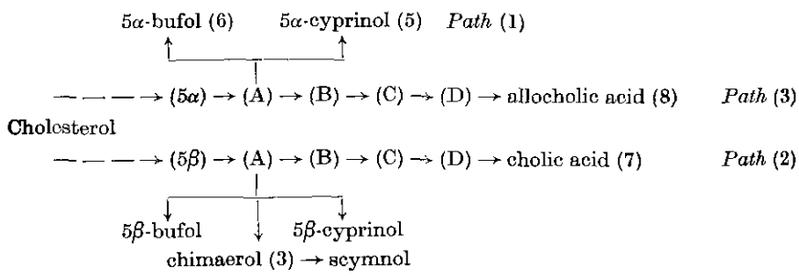


FIG. 1. Possible pathways for biosynthesis of bile alcohols and bile acids

Path (1) is mainly followed in cyprinid fishes, but it is evident that Path (2) is also activated in these forms. In almost all other teleosts, Path (2) is the only one detected.

The evidence for saying that Path (1) is primitive is that 5α-cyprinol is a bile alcohol of unquestionably primitive osteichthyeans. There is also a progression amongst the Ostariophysi, for the silurids seem more advanced than the cyprinids, some having a little 5α-cyprinol but most containing only cholic acid.

Thus, there is a strong case for holding that, in their bile salts, the cyprinids and silurids are primitive teleosts and that the cyprinids are the more primitive of these two groups.

A similar view was at one time held by some taxonomists; for instance Newman (1939) says: "The carps are often cited as being the most generalized or most nearly ancestral of all surviving teleosts". In their recent classification, Greenwood *et al.* (1966) very tentatively assign the Ostariophysi a position close to the origin of their Division III, but above the salmonids. No 5α-cyprinol has yet been found in salmonoid bile: two species (*Plecoglossus altevelis* and *Oncorhynchus rhodurus*) have been very carefully examined (references in Haslewood, 1967b) and I have found only taurocholate and taurochenodeoxycholate in *Coregonus clupeoides*.

The separation of teleosts into separate possibly polyphyletic groups by Greenwood and his colleagues finds some support from bile salt studies, for traces of 5β-cyprinol (3α,7α,12α,26,27-pentahydroxy-5β-cholestane) have been found in the bile of certain eels. 5β-Cyprinol could arise by a single step from the 5β epimer of (A). It is one of the bile alcohols in sturgeons.

The holostean *Amia calva* (bowfin) contains only taurocholate.

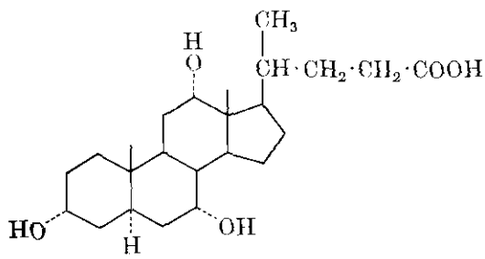
It seems very probable that bile salt studies will be able to contribute to our understanding of teleost taxonomy.

AMPHIBIA

Amphibian bile salts are complex, containing both bile alcohols and bile acids. 5α-Cyprinol (5) is present in *Megalobatrachus japonicus* (Cryptobranchidae), 5α-bufol (6) in *Diemyctylus pyrrhogaster* (Salamandridae) and the 5β-epimer of (6), 5β-bufol (3α,7α,12α,25ξ,26(or 27)-pentahydroxy-5β-cholestane) seems to be characteristic of species of *Bufo*. As mentioned above, both 5α-cyprinol and 5α-bufol are present in the dipnoan *Neoceratodus* (it is not determined whether 5α-bufol occurs also in *Lepidosiren* or *Protopterus*) and an animal with bile salts like this would be an acceptable amphibian ancestor. Indeed, there is a

striking resemblance between *Neoceratodus* and *Diemyctylus* bile, for the newt also contains 5α-cyprinol (Hoshita *et al.*, 1967). If amphibians inherited genes for the production of 25-hydroxylating enzymes from fishes, as seems likely, *Latimeria* is apparently excluded as a forbear, for 5α-bufol has not been demonstrated in this coelacanth. The bufonids might have gone over to the 5β series (Fig. 1), whilst still retaining the bufol type of bile alcohol; they seem to possess no ability to proceed along Path (2) further than stage (C) and do not make cholic acid.

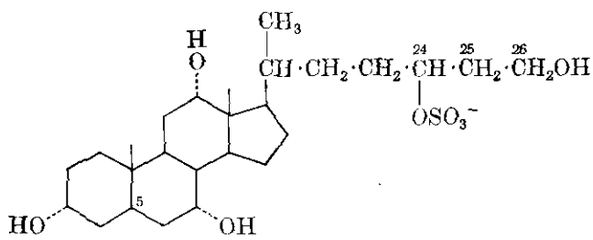
Megalobatrachus is also using Path (3) (Fig. 1) and makes allocholic acid (3α,7α,12α-trihydroxy-5α-cholanoic acid (8)); this is also true of *Salamandra salamandra* and of many lizards.



(8) Allocholic acid

The Ranidae (frogs) have remarkable bile salts. Small amounts of cholic acid are found and also of the 5β-epimer of substance (C) (Path (2), Fig. 1); however in the species examined thoroughly (*Rana catesbiana, R. nigromaculata, R. temporaria*) bile alcohols are prominent; in *R. nigromaculata* 5β-cyprinol is present, but *R. catesbiana* and *R. temporaria* have respectively the 5β- and 5α-ranols (3α,7α,12α,24ξ,26-pentahydroxy-5α- or 5β- 27-norcholestanes) as their C-24 sulphates (9).

The biosynthetic origin of these, in relation to Fig. 1, awaits clarification. It is, however, abundantly clear that the ranids are heterogeneous

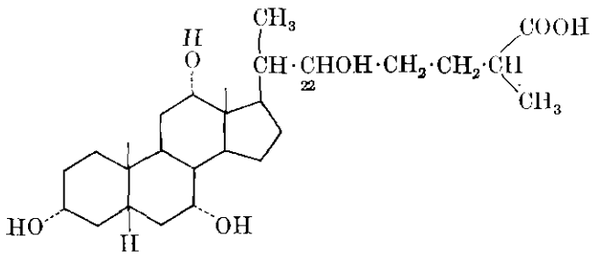


(9) Ranol sulphates

in *their bile salts* and it would not be unreasonable to suppose that the ranols are biochemically more advanced than 5β-cyprinol. There is clearly no close relationship between *Rana* and *Bufo* in bile salts and *Bufo* has the more primitive type.

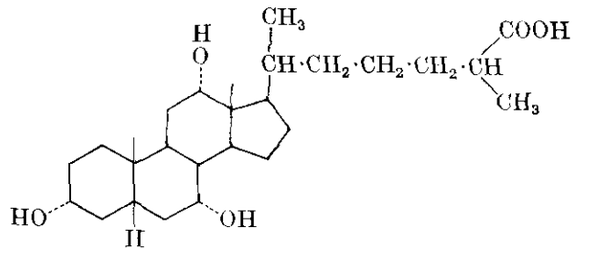
REPTILES

The bile salt evidence accords with the idea (Partington, 1962) that the chelonians perhaps arose from a stock (Diadectomorpha) distinguishable from the Captorhinomorpha from the earliest fossil records, for no bile acid has yet been found common to chelonians and other tetrapods. An acid, 3α,7α,12α,22ξ-tetrahydroxy-5β-cholestan-26(or 27)-oic acid (10) identified in turtle and tortoise bile can be regarded as both primitive and specialized.



(10) 3α,7α,12α,22ξ-Tetrahydroxy-5β-cholestan-26(or 27)-oic acid

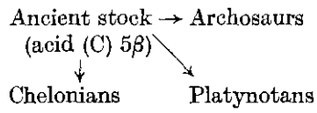
The acid (10) could arise by 22-hydroxylation of the 5β epimer of compound (C) (Fig. 1), i.e. 3α,7α,12α-trihydroxy-5β-cholestan-26-oic acid (11).



(11) 3α,7α,12α-Trihydroxy-5β-cholestan-26-oic acid

The acid (11) itself is found as a chief constituent of crocodilian bile and may possibly represent the primitive C_{27},5β bile acid from which

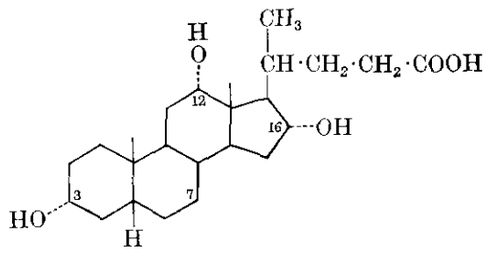
arose not only the chelonian bile acid (10) but also the $3\alpha,7\alpha,12\alpha,24\xi$-tetrahydroxy-$5\beta$-cholestan-26(or 27)-oic acid ((D) of Fig. 1) of *Varanus* and *Heloderma*. Thus, the bile acid evidence accords with the picture



On this view, snake ancestors could have been precursors of the platynotans, for snakes contain almost entirely C_{24},5β acids derivable (Fig. 1) from acid (C) (5β).

In other words, snake ancestors perhaps had already taken the 5β pathway, in contrast to the precursors of most of the lizards examined (except *Varanus* and *Heloderma*), which have $C_{24}5\alpha$, bile acids.

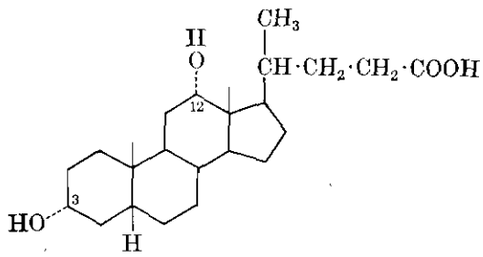
There are certain secondary bile acids in snakes which may eventually be useful for classification. Pythocholic acid ($3\alpha,12\alpha,16\alpha$-trihydroxy-$5\beta$-cholanoic acid (12) is derived by 16α-hydroxylation of deoxycholic



(12) Pythocholic acid

acid ($3\alpha,12\alpha$-dihydroxy-5β-cholanoic acid (13) which in turn is made from cholic acid (7) by intestinal microflora (Bergström *et al.*, 1960).

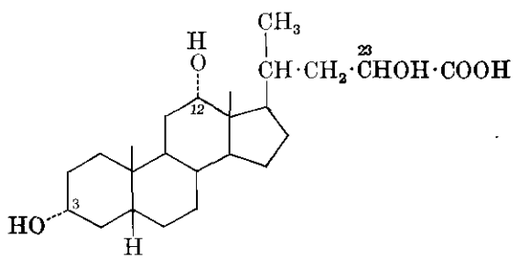
Pythocholic acid is present in at least 12 species of Boidae and



(13) Deoxycholic acid

probably also in *Cylindrophis rufus* (Aniliidae): thus 16α-hydroxylation seems to be a primitive snake character. It is remarkable that it was not found in the boid *Corallus enhydris*, although of course this may have been due to the absence of micro-organisms making deoxycholic acid in this animal.

Perhaps, however, the "patchy" occurrence of a biochemical character like this is of taxonomic significance, for the occurrence of C-23 hydroxylation in several more advanced snakes appears at present to make limited systematic sense. Bitocholic acid (3α,12α,23ξ-trihydroxy-5β-cholanoic acid (14) may, like pythocholic acid, be surmised (Underwood, 1967) also to be a secondary bile acid, made from deoxycholic acid (13).



(14) Bitocholic acid

Bitocholic acid and C-23 hydroxylated cholic acid (3α,7α,12α,23ξ-tetrahydroxy-5β-cholanoic acid) were first isolated from *Bitis* species and have since been detected in most of the Viperinae examined. This C-23 hydroxylation, however, is absent from *Cerastes cerastes* and *Echis carinatus* (Viperinae) and from all the Crotalinae examined (11 species). It also occurs in some colubrid snakes.

There is no instance of two species of the same genus having different bile acids, but the occurrence of C-23 hydroxylation seems to transgress some of the wider boundaries drawn by Underwood (1967) and other snake systematists in a manner that strongly invites further (especially biochemical) investigation.

BIRDS

So little is known about bird bile salts that speculation about them is hardly justified. No specifically "reptilian" bile acids have been found, except perhaps the unusually large amounts of allocholic acid (8) in two species of penguins. Allocholic acid, as already mentioned, is the chief bile acid in many lizards. The presence of this substance (if not a

dietary artifact in penguins) might conceivably be a primitive biochemical character but could not be classed as such on the present meagre information about bird bile.

MAMMALS

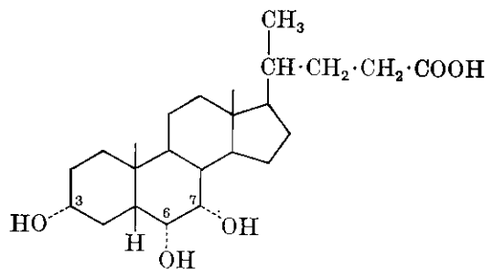
Nothing has been found in monotremes that would suggest any biochemical bile salt difference between these animals and the remaining mammals. Glycine conjugates are absent from monotreme bile and are also apparently missing or minimal in marsupials. They first make their appearance in eutherian mammals. In these, the enzymes for the reaction:

$$R.CO.S.\underline{CoA} + H_3\overset{+}{N}.CH_2.COO^- \rightarrow R.CO.NH.CH_2.COO^- + \underline{CoA}.SH + H^+$$

Bile acid CoA derivative — Glycine — Glycine conjugate — CoA

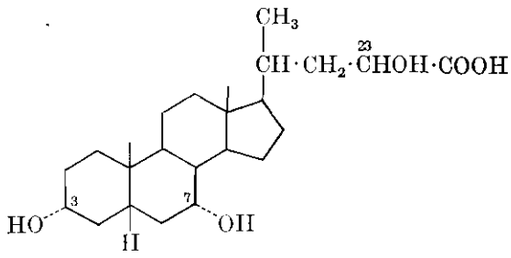
are active and functional in many species, but apparently dormant or missing in others. Further biochemical study might be useful in relating glycine conjugation to mammalian systematics, especially in rodents. As an example, the chief bile salt difference between *Myocastor coypus* and *Thyronomys swindereanus* is the apparent absence of glycine conjugates in the latter animal.

Unique bile acids are found in some eutherians. For example, pigs (Suidae) all have 6α-hydroxylated bile acids, like hyocholic acid (3α,6α,7α-trihydroxy-5β-cholanoic acid (15)).



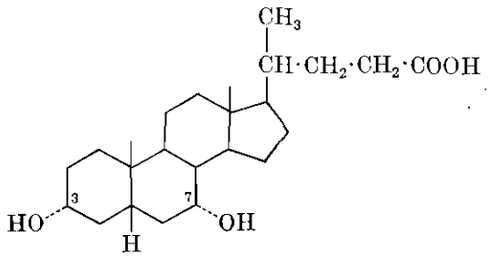
(15) Hyocholic acid

All the Pinnipedia apparently have phocaecholic acid (3α,7α,23ξ-trihydroxy-5β-cholanoic acid (16)) and Murinae have small amounts of 6β-hydroxylated-5β-cholanoic acids.



(16) Phocaecholic acid

All these acids can be regarded as derived from chenodeoxycholic acid (3α,7α-dihydroxy-5β-cholanoic acid (17))



(17) Chenodeoxycholic acid

and have been in some cases shown to be made from this substance. Now chenodeoxycholic acid is the second principal C_{24} bile acid and its biosynthesis probably differs from that of cholic acid (Danielsson and Tchen, 1968). Moreover, there is a clear tendency to dihydroxy (rather than trihydroxy) C_{24} bile acids in several animal groups, especially mammals having herbivorous dietary habits. Thus, it is possible that acids like hyocholic and phocaecholic might have arisen by rehydroxylation of chenodeoxycholic acid in animals requiring trihydroxy bile acids in place of chenodeoxycholic acid. A taxonomic implication would be that the ancestors of suids and pinnipeds included more strictly vegetarian forms. It would be interesting to have the views of mammalian systematists on such a question.

CONCLUSION

When the primary bile salt relationships are set out stepwise as in Fig. 1, one is tempted to think of them as mediated by enzymic sequences of the operon type in the Jacob-Monod sense: i.e. controllable

by regulator genes. If this were so, bile salt evolution could consist of the "switching on" of whole gene sequences leading to physiologically improved bile salts and the repression and eventual degradation of more primitive gene successions. Such ideas may be helpful in assessing the taxonomic importance of the differences discussed in this communication.

ACKNOWLEDGMENTS

The author thanks Dr. A. D. Cross and Dr. L. Tökés of the Institute of Steroid Chemistry, Syntex Research, Palo Alto, California, for nuclear magnetic resonance and mass spectral observations on chimaerol. He also thanks Dr. T. Briggs for *Latimeria* bile salts and Dr. P. S. Maitland for *Coregonus* bile.

REFERENCES

Anderson, I. G., Haslewood, G. A. D., Cross, A. D. and Tökés, L. (1967). New evidence for the structure of myxinol. *Biochem. J.* **104**, 1061–63.

Bergström, S., Danielsson, H. and Kazuno, T. (1960). Bile acids and steroids 98. The metabolism of bile acids in python and constrictor snakes. *J. biol. Chem.* **235**, 983–88.

Danielsson, H. and Tchen, T. T. (1968). Steroid Metabolism. *In* Greenberg, D. M. (ed.) "Metabolic Pathways". Vol. 2 (3rd ed.). 117–68 Academic Press, New York.

Greenwood, P. H., Rosen, D. E., Weitzman, S. H. and Myers, G. S. (1966). Phyletic studies of teleostean fishes, with a provisional classification of living forms. *Bull. Am. Mus. nat. Hist.* **131**, 339–456.

Haslewood, G. A. D. (1967a). "Bile Salts." Methuen, London.

Haslewood, G. A. D. (1967b). Bile salt evolution. *J. Lipid Res.* **8**, 535–50.

Haslewood, G. A. D. and Tökés, L. (1968). Petromyzonol sulphate, a bile salt of lampreys. *Biochem. J.* **107**, 6P.

Heintz, A. (1962). Phylogenetic aspects of myxinoids. *In* Brodal A. and Fänge, R. (eds.) "The Biology of Myxine" 16. Universitetsforlaget, Oslo.

Hoshita, T., Hirofugi, S., Nakagawa, T. and Kazuno, T. (1967). Stero-bile acids and bile alcohols. 96. Studies on the bile salts of the newt and synthesis of 5α-cholestane-3α,7α,12α,25,26-pentol (5α-bufol). *J. Biochem., Tokyo* **62**, 62–6.

Newman, H. H. (1939). "The Phylum Chordata" 187–8. McMillan, New York.

Partington, F. R. (1962). Les relations des Cotylosaures diadactemorphes. *Colloques internationaux du C.N.R.S.* No. 104. "Problèmes actuels de paléontologie (Evolution des Vertébrés)" 175–85. Paris.

Patterson, C. (1965). The phylogeny of the chimaeroids. *Phil. Trans. R. Soc.* Series B. **249**, 212.

Underwood, G. (1967). "A Contribution to the Classification of Snakes." The British Museum (Natural History), London.

16. The Use of Secondary Chemical Characters in the Systematics of Higher Plants

J. B. HARBORNE

Phytochemical Unit, The Hartley Botanical Laboratories, University of Liverpool, England

ABSTRACT

It is shown that secondary plant constituents have an important place in the application of chemistry to higher plant systematics. Plant to plant variation is discussed and the valne of replacement over presence/absence characters is stressed. The three main methods of chromatographic analysis used in the screening of plants for secondary constituents are described briefly. The use of chemistry in the identification of species and of species hybrids is outlined. An example is given of the identification of a *Primula* plant incorrectly assigned on morphological grounds to the *P. rosea* group but shown by paper chromatography to be identical with *P. warshenewskiana*. The verification, through a study of the leaf alkaloids, of an unusual *Solanum* tetraploid hybrid which arose by double chromosome reduction from the hexaploid *S. demissum* and the tetraploid *S. stoloniferum* is described.

The use of secondary constituents at higher levels of plant classification is also discussed. A survey of the flavonoid pigments in the family Gesneriaceae is shown to support the recent classification proposed by Burtt on morphological grounds. Differences in pigment type at the subfamily level indicate a correlation between chemistry and plant geography, which has been traced to the different pollinating mechanisms in the contrasting habitats of the Old and New World gesnerads. Some results of a chemical survey of the tribe Caucalineae of the Umbelliferae are presented and it is shown that examination of the polyphenols reinforces the conclusions from protein studies about generic delimitations in the group.

INTRODUCTION

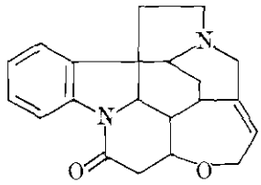
Secondary plant constituents

Plants, unlike animals, accumulate in their cells a considerable range of low molecular weight substances, the majority of which have no

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 173–191.

distinct function in primary metabolism. These substances, the so-called secondary constituents, are often present in considerable quantities and are sometimes regarded, incorrectly in my view, as waste products of plant metabolism. They range from the toxic alkaloids and the odoriferous essential oils to the cathartic anthraquinones and the coloured anthocyanins and flavones of flower petals. These substances, because they are not directly involved in the essential metabolic activity of the plant, vary immensely both qualitatively and quantitatively in their distribution and offer an excellent range of characters for use in a chemical approach to plant systematics.

The rapid development of biochemical systematics in the last ten years (see Swain, 1963, 1966) has owed much of its impetus to the study of these secondary constituents. Compared with the large molecules of plants, they have several advantages. For example, there is rarely any ambiguity about their structures. Many are chemically complex, e.g. the alkaloid strychnine (Formula 1), but the labour involved in



(1) Strychnine

structural identification using the modern physical methods now available is considerably less than that required for elucidating the amino acid sequence of the simplest protein. Their biosynthesis requires a range of enzymes above the normal complement for essential metabolism, and to study the distribution of strychnine in *Strychnos* species is in effect to make an enzymic comparison between species. Thus, any conflict between the use of large and small molecules in higher plant systematics is largely illusory.

There are several different ways of assessing the results of surveying plants for secondary constituents. One can regard it simply as an exercise in comparative biochemistry, a perfectly valid approach. One can also make use of the information so gained for unravelling biosynthetic pathways in plants. Again, one may regard the vast body of chemical information available about plants from the phyletic point-of-view and construct chemo-evolutionary trees from such data. I wish here to concentrate on a more down-to-earth approach; to indicate the value of surveying plants for secondary constituents to the taxonomist concerned with plant identification and with establishing taxonomic

hierarchies among restricted plant groups. I hope to show that chemistry can provide significant systematic information not available from any other approach and that chemistry can solve, on occasion, a systematic problem that has failed to yield to cytological, anatomical or morphological techniques.

Before giving some examples of the use of chemistry in plant systematics, I must say a few words about how secondary constituents vary from plant to plant and about the techniques required for determining their distribution patterns.

Plant to plant variation

The occurrence of a chemical in a plant can be assessed by the chemotaxonomist in three ways: (a) qualitatively, i.e. as a presence/absence character; (b) quantitatively; and (c) biosynthetically, as one character replacing another. The majority of surveys are carried out at the qualitative level and one obtains data in the form shown in Table 1. Quantitative studies become necessary when analysing plants for essential oils and leaf waxes, since the same spectrum of terpenes and

TABLE 1. An example of presence/absence characters: the flavonoids of Pine heartwoods

Subgenus and subsection (no. of species surveyed in parenthesis)	Flavones			Flavanones			Flavanonols	
	1	2	3	4	5	6	7	8
HAPLOXYLON								
Cembrae (3)	+	+	−	+	+	−	+	−
Flexiles (1)	+	+	−	+	+	−	+	−
Strobi (8)	+	(+)	(+)	+	(+)	(+)	+	(+)
Cembroides (1)	+	+	−	+	+	−	+	−
Gerardianae (2)	−	−	−	+	+	(+)	+	+
Balfourianae (2)	+	+	−	+	+	−	+	−
DIPLOXYLON								
Leiophyllae (2)	−	−	−	+	−	−	+	−
Longifoliae (2)	−	−	−	+	−	−	+	−
Pineae (1)	−	−	−	+	−	−	+	−
Lariciones (7)	−	−	−	+	−	−	(+)	−
Australes (9)	−	−	−	+	−	−	+	−
Insignes (12)	−	−	−	+	−	−	+	−
Macropae (2)	−	−	−	+	−	−	+	−

Key: 1, chrysin; 2, tectochrysin; 3, strobochrysin; 4, pinocembrin; 5, pinostrobin; 6, strobopinin; 7, pinobanksin; 8, strobobanksin. +, uniformly present; (+) not present in all species; −, absent. Data adapted from Erdtman (1956). Subsectional names are derived from type species within the subsection, e.g. Cembrae after *P. cembra* L., Flexiles after *P. flexilis* Jam., etc.

hydrocarbons tends to be present uniformly throughout a given group of plants. Concentrations may vary significantly from species to species and these are usually given as a percentage of the total volatiles (Table 2). The systematic assessment of quantitative variation is

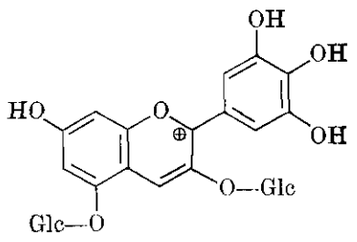
TABLE 2. Turpentines of some Mediterranean pines

SPECIES	TERPENE COMPOSITION						
	α-Pinene	Camphene	β-Pinene	3-Carene	Myrcene	Limonene	β-Phellandrene
P. eldarica	56·0	0·5	4·0	29·0	1·5	0·5	1·0
P. stankewiczii	57·0	trace	6·5	26·5	1·0	0·5	trace
P. brutia	79·5	1·0	6·0	9·0	trace	1·5	0·5
P. pityusa	69·8	—	—	23·6	—	—	—
P. halepensis	95·0	—	—	—	—	—	—
P. sylvestris	60·6	—	17·4	16·9	—	—	—

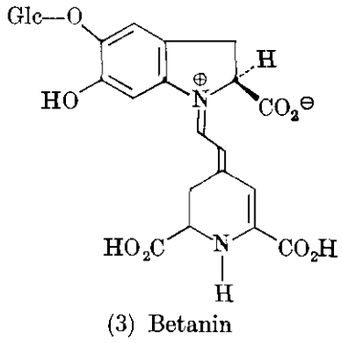
Figures are given as % of total turpentine. The first four species are all considered by some taxonomists to be synonymous with *P. halepensis* but the terpene results indicate that they are chemically distinct. Data from Mirov, Zavarin and Snajberk (1966).

difficult and much thought still needs to be given to the analysis of such data.

Presence/absence characters have certain disadvantages: for example, it is not usually possible to prove that a chemical is completely absent from a plant; one can only write down "not detectable at a given concentration by the technique used". This difficulty vanishes when one chemical character occurring in some species is uniformly replaced by a related character in other species and one is then in a particularly strong position to apply the results to systematics. Unfortunately, few such unambiguous situations have so far emerged during chemotaxonomic studies. The best known example is that of the replacement of the widespread purple anthocyanins (e.g. Formula 2) by the similarly coloured nitrogenous betacyanins (e.g. Formula 3) in just one order of



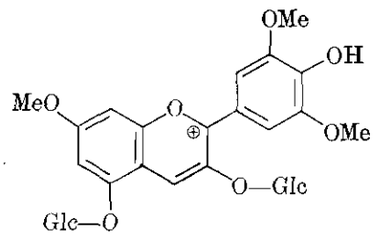
(2) Delphin



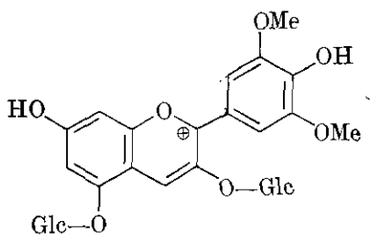
(3) Betanin

higher plants, the Centrospermae. The two classes of pigment have never yet been found to co-occur, with one possible exception, in the same plant. Although they differ considerably in biosynthetic origin, their visual effect is very similar. The situation is interesting taxonomically since the presence of betacyanins in two families once separated from this order, the Cactaceae and the Didieraceae, has been used to confirm their inclusion in the order. There is, however, one apparently good centrospermous family, the Caryophyllaceae, which retains anthocyanin pigmentation and there is thus chemical justification for re-examining its affinities.

A simpler example of a replacement character is that of the rare anthocyanin pigment, hirsutin (Formula 4), which replaces the much more common anthocyanin, malvin (Formula 5), in several subgenera



(4) Hirsutin



(5) Malvin

of the genus *Primula*. Hirsutin only differs from malvin in having an extra methyl group and is presumably synthesized from it by a single enzyme-catalysed methylation. My recent surveys (Table 3) show that while hirsutin occurs characteristically in the subgenera *Auriculastrum*, *Primula*, *Craibia* and *Aleuritia*, it is replaced by the simpler malvin in the subgenus *Auganthus*, which includes three popular ornamental species *P. obconica*, *P. sinensis* and *P. malacoides*. I am following here Wendelbo's recent reclassification (1961) of *Primula* and it is pleasing to find the chemistry of the flower pigments agrees for the most part with his subgeneric revisions.

The chemical analysis of plants is complicated by the presence of "chemical races" in some species, by genetic variation in others and environmental effects in yet others. I have no time to dwell in detail on these complications but it is true to say that chemical characters by and large are no less stable than most so-called 'reliable" biological characters. Just as leaf shape, a very plastic character in some circumstances is universally used in taxonomic assessments, so anthocyanin pigmentation, subject to considerable genetic variation in some

TABLE 3. An example of a replacement character—flavonoid pigments in the genus *Primula*

Subgenus	No. of spp. examined	Anthocyanins		Yellow pigments	
		Common	7-Methylated	Carotenoid	Flavonoid
Sphondylia	1	—	—	present	—
Auriculastrum	5	—	hirsutin (5)	—	—
Primula	4	—	hirsutin (1)	traces	gossypetin (4)
Auganthus	5	malvin (5)	—	—	—
Craibia	3	—	hirsutin (3)	—	—
Aleuritia	29	malvin (1) cyanin (3)	hirsutin (13) rosinin (2)	present (7)	gossypetin (3) (in section Sikkimensis only)

Classification after Wendelbo (1961); chemical results from unpublished work of the author. Numbers in parentheses indicate number of species with the character indicated, a dash indicating that the character is absent. Hirsutin is, however, often accompanied by common anthocyanins as trace constituents.

ornamental plants, may be used with confidence as a chemical marker in most plant groups.

Methods of chemical analysis

The screening of plants for secondary constituents—once a very laborious and esoteric art—is now a simple, speedy process, due to the advent of chromatography. Present-day techniques are well within the capacities of biologists, although of course, whenever new substances are encountered, it is as well that a chemist should be at hand to isolate and identify them. There are three main types of chromatography—paper, thin-layer and gas–liquid—and I shall give brief examples in turn of the results obtained from each technique.

PAPER CHROMATOGRAPHY

Paper chromatography—the technique which twenty years ago won a Nobel Prize for its inventors—is now being steadily replaced by the more versatile thin-layer chromatography for many separations. It is however still the technique of choice for examining water-soluble plant constituents such as sugars and amino acids, and its use is illustrated here with reference to the flavonoids and related phenolic constituents. Figure 1 shows two-dimensional paper chromatograms of leaf and petal extracts of two common British plants, the cowslip *P. veris* and the primrose *P. vulgaris*. The phenolic compounds are detected by examination of the paper under ultra-violet light where they appear as dull or

fluorescent spots in a range of colours from black, brown, yellow to green and white. As can be seen, there is a considerable number of phenolic constituents in these *Primulas*, some of which occur in both species and others of which are species-specific. Chromatograms so produced remain unchanged almost indefinitely and could be stored in herbaria for reference purposes alongside plant specimens.

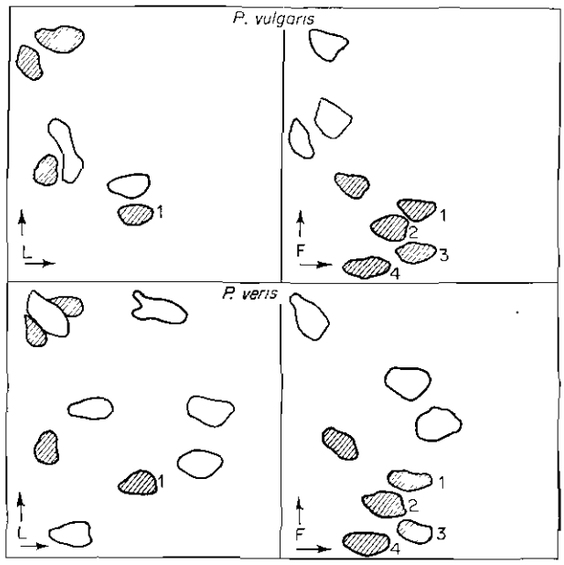


FIG. 1. Two-dimensional paper chromatograms of *Primula* species. Key: 1 = kaempferol 3-gentiotrioside; 2 = quercetin 3-gentiotrioside; 3,4 = gossypetin glycosides. Solvents: in upward direction, butanol-acetic acid-water (4:1:5) from left to right 5% aqueous acetic acid. L = *leaf extract*, F = *flower* extract. Hatched spots: compounds common to both species. Open spots: species-specific compounds.

THIN-LAYER CHROMATOGRAPHY

Thin-layer chromatography is really only a modification of paper chromatography and is more versatile simply because a range of adsorbents besides cellulose may be used. The adsorbent is spread as a thin layer on a clean glass plate and the plate is then developed in just the same way as a paper chromatogram. T.L.C. is the most important technique for separating substances such as alkaloids, quinones and coumarins. Typical one-dimensional separations on silica gel of the anthraquinones in aloe samples from different parts of the world are shown in Fig. 2 (see Hörhammer *et al.*, 1965).

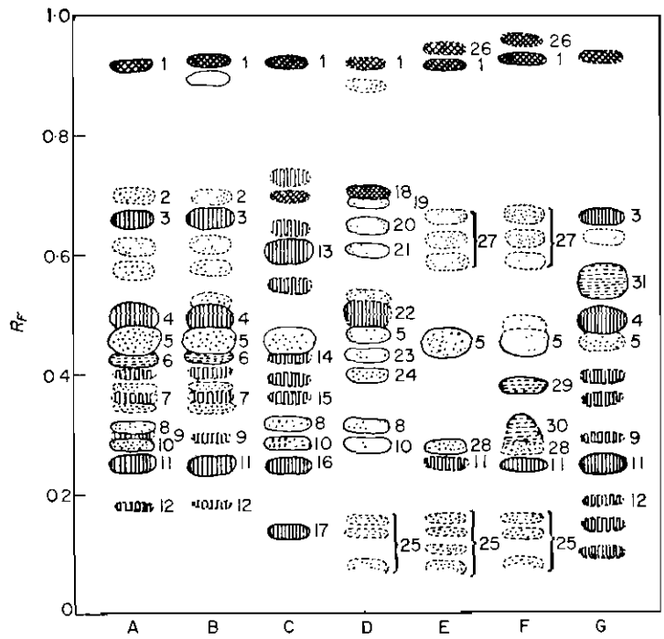


FIG. 2. Thin layer chromatogram of anthraquinones in aloes (Hörhammer *et al.*, 1965). Adsorbent: silica gel. Solvent: ethyl acetate–methanol–water (100:17:14). Plate sprayed with methanolic alkali and examined under ultra-violet light. A, B = cape aloe (*Aloe* spp. and hybrids); C = East African aloe; D = Socotra aloe (*A. perryi* Baker); E = Indian aloe (*Aloe* spp.); F = West Indian aloe (*A. barbadensis* Miller); G = Natal aloe (*Aloe* spp.). Various shadings indicate colours in u.v. light as follows:

= yellow; = red;

= blue green; = brown violet.

GAS–LIQUID CHROMATOGRAPHY

Gas–liquid chromatography is the technique of choice when dealing with the volatile plant constituents and the method has the advantage that the results are given in both qualitative and quantitative terms. The volatiles are separated on an inert column coated with a silicone at a temperature of 150° or so and appear as a series of peaks on a chart, the heights of which are related to the concentrations. The gas-chromatograph traces of the essential oils of three species of catmint

(Fig. 3) indicate the large amount of information produced by this one technique. Each of the three species can immediately be distinguished by a major oil character: citronellol in *Nepeta citriodora*, nepetalactone in *N. cataria* and epinepetalactone in *N. mussini* (see Regnier *et al.* 1967).

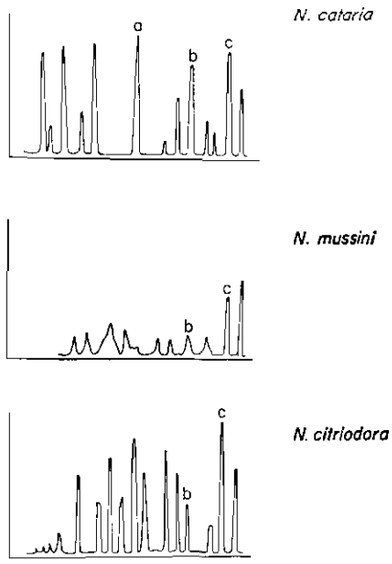


FIG. 3. Gas chromatograph traces of the essential oils from *Nepeta* species. a, camphor; b, carophyllene; c, nepetalactone (Regnier *et al.*, 1967). Separation on an O5–138 polyethyl other column. Ordinate: recorder response; abscissa; temperature and time, increasing from left to right.

(Reproduced by kind permission of Pergamon Press).

CHEMISTRY AND PLANT IDENTIFICATION

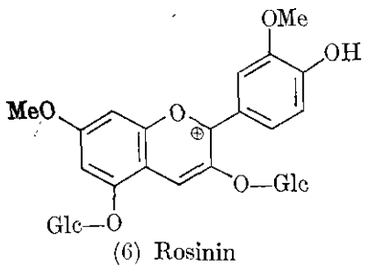
Species identifications

The identification of higher plants solely on morphological characters is not always straightforward; every system of classification has its "difficult" groups. While chemistry can never compete with the elaborate morphological keys used in plant identification, it might profitably be used at the species level in difficult cases. Chromatographic tests, which can often be carried on a fragment of dried leaf and give a result within an hour, might eventually become part of the taxonomist's armoury in some plant groups.

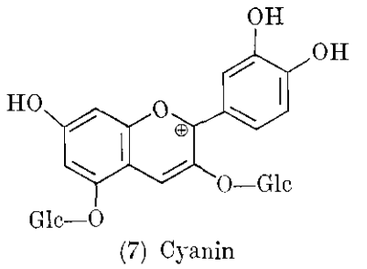
Special difficulties attend the identification of drug plants, because only a portion of the plant is usually available for study. Here, chemistry

has already taken over from anatomical studies and thin-layer chromatography is widely used in pharmacies, especially in Germany, for checking the authenticity of herbal supplies and for looking for contaminating materials in root and leaf samples. The correct identification of plant varieties is a matter of some economic importance, especially now that patent laws are being introduced in England. This is often tricky, since so few clear-cut morphological characters are available. Chemical variation in secondary constituents occurs in most cultivated crops (see, e.g., Rowlands and Corner, 1963) and such differences may well be sought when carrying out cultivar identifications.

Phytochemical experience indicates that one could find in every wild plant species one or more species-specific constituents, which could be used potentially for purpose of identification. Let me quote one example from my own survey of the genus *Primula*, a taxonomically confusing group of over 500 species. In surveying the collection of plants growing in the University of Liverpool Botanic Gardens at Neston, I came across a plant labelled "*P. aff. rosea*", which had been collected on one of the Furze Himalayan expeditions. It had the characteristic rose pink petal colour of *P. rosea* and was so similar morphologically that there seemed no reason to doubt its affinities. I had, however, earlier found a very rare anthocyanin, rosinin (Formula 6), in *P. rosea* which was quite



(6) Rosinin



(7) Cyanin

characteristic of the species (Harborne, 1960) and on examining the *P. aff. rosea* plant, I immediately became suspicious, since it lacked this pigment, having the much more common anthocyanin, cyanin (Formula 7), instead. Pink flower colour is relatively rare in the genus, but there is another species *P. warshenewskiana* with this colour in the same section. Examination of the chromatographic patterns of petal extracts of *P. aff. rosea* and *P. warshenewskiana* (Fig. 4) showed them to be identical, both having cyanin and six other flavonoid components in common. By contrast, *P. rosea* is very different, having only three of the seven constituents of *P. warshenewskiana* and in addition three further compounds, including rosinin. This chemical comparison rules

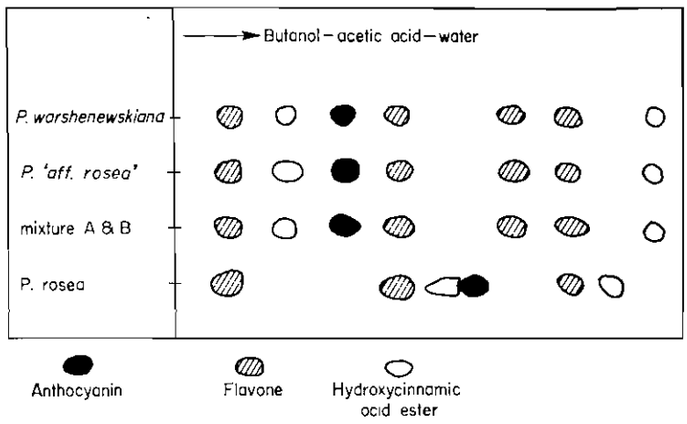


FIG. 4. Identification of *Primula aff. rosea* by chemistry of its petal pigments.

out the possibility of simple genetic variation and indicates that the Furze material is identical to *P. warshenewskiana*; it certainly shows quite unambiguously that it is not a form of *P. rosea*.

Hybrid identification

The identification of plant hybrids by means of morphology and anatomy is sometimes very difficult and the fact that hybridization readily occurs in a particular plant genus considerably confuses any taxonomic treatment of that group. Cytology has helped considerably in this situation in distinguishing hybrids from true species and evidence is accumulating that chemistry can play an equally useful role in hybrid studies.

Hybrid plants generally contain, at least in their leaves, the chemical constituents of both parents, and chromatographic studies of a hybrid extract will show a pattern of spots more or less identical to a synthetic mixture of extracts of the two parents. Such chromatographic studies of leaf phenolics have been used extensively for identifying hybrids among natural populations of the legume genus *Baptisia* by Alston, Turner and their co-workers in Texas. These authors claim (Alston and Turner, 1963) that the chemical studies succeeded in proving the parentage of some hybrids where morphological, anatomical and cytological measurements had completely failed.

Occasionally, completely new chemical compounds are found in plant hybrids. Nybom (personal communication), for example, in examining

the anthocyanins of currants, found two new anthocyanins in the berries of a cross between *Ribes sativum* and *R. nigrum* (Fig. 5). The two "hybrid" pigments are clearly related in structure to those of the parents and their presence is very useful from the point-of-view of confirming that hybridization had occurred.

My own example here is from a study by Marks *et al.* (1965) of an unusual *Solanum* hybrid, produced by crossing two South American

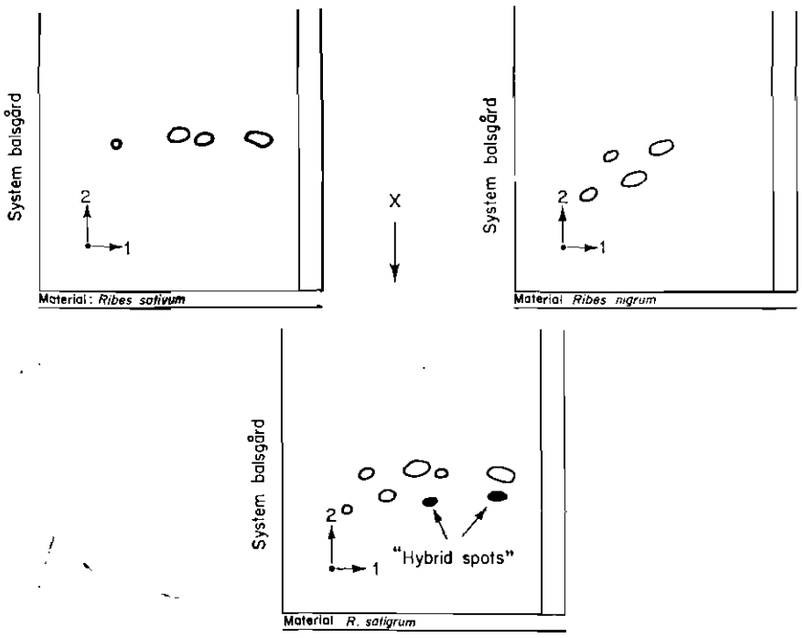


FIG. 5. Anthocyanins in the fruit of a *Ribes* hybrid and of the parent species (F. Nilsson and N. Nybom; unpublished data).

tuber bearing species, *S. demissum* and *S. stoloniferum*. It was of importance to establish the hybrid nature, because the cytology was unusual; indeed, this hybrid proved to be the first fully documented case of double chromosome reduction in higher plants. The plant was a tetraploid produced from crossing a tetraploid and a hexaploid and morphologically it was close to the hexaploid parent and showed no obvious characters of the tetraploid parent. It could conceivably have arisen by male or female parthenogenesis. However, chemical studies of the leaf alkaloids (Fig. 6) showed quite incontestably that it was a

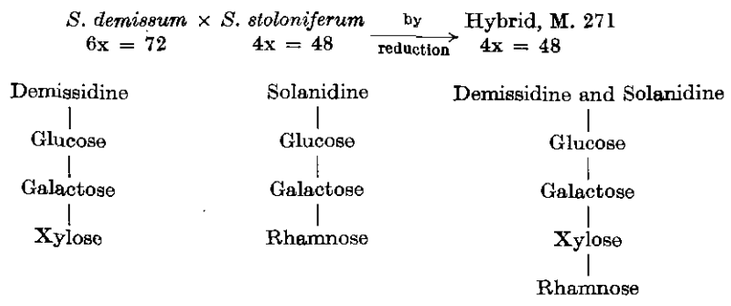


FIG. 6. Leaf alkaloids of *Solanum* parents and hybrid.
(Note: alkaloid glycosides are of more complex structure than shown here.)

true hybrid, having alkaloid-bases and combined sugars coming from both parents. In fact, the plant contained a new "hybrid" alkaloid, the structure of which has not yet been determined. It is worth pointing out here that the study of the phenolics in flowers and leaves in this case, although of value in showing that other related *Solanum* species were not involved in the cross, failed to indicate its hybrid nature. Thus in the general situation, it is necessary to examine more than one class of chemical constituent to be sure of tracking down a suspected hybrid plant.

SECONDARY CHEMICAL CONSTITUENTS AND PLANT CLASSIFICATION

Besides providing a means of naming plants, higher plant systematics is concerned with expressing plant relationships in classificatory terms and fitting taxa into a hierarchic system, usually constructed on an evolutionary basis. Here, chemistry obviously has a place and has indeed contributed, admittedly to a very minor extent, to the well-known systems of Engler and Prantl and Bentham and Hooker. Taxonomists from classical times have taken into account the scent of a plant and the colour of its petals. Now that we are no longer dependent on sniffing a plant to determine its essential oil content or on crude colour tests to identify its flower pigments, we are in a much stronger position to apply chemistry to systematic ends. Certainly, a unique chemical feature should merit the same importance as an odd chromosome number or an unusual leaf shape in determining plant relationships.

There is understandably some resistance from taxonomists to using chemical characters and indeed, one can point to many examples (see

Hegnauer, 1966) where, say, alkaloid occurrences* obscure rather than illuminate plant relationships. However, many of the difficulties are either due to incomplete knowledge of alkaloid distribution and biosynthesis or due to attempts to draw chemical conclusions at the higher taxonomic levels. More will be learnt at the present stage by restricting phytochemical surveys and the discussion of results to the smaller plant groups of the genus or the tribe.

Here, chemists have already shown that chemical characters can correlate in generic surveys most satisfactorily with one or more biological characters. One of the first such correlations to be observed was that of Erdtman (1956), in his now classic survey of the heartwood constituents of over 50 pine species. He found (see Table 1) that the flavonoid pattern was exactly correlated with the morphological division of the genus *Pinus* into two subgenera Haploxylon and Diploxylon, based on the number of vascular bundles in the pine needles. The flavonoid pattern in the more primitive Haploxylon series is complex, while that in the more advanced Diploxylon is simpler. I wish to discuss two more recent examples taken from my own work to illustrate the type of systematic data oue may obtain from the chemical standpoint.

Pigments of the Gesneriaceae

One situation where chemistry can aid taxonomy is to confirm or otherwise the reclassification of a particular plant group. An example of this is in the family Gesneriaceae, which was divided into two subfamilies by Fritsch in 1893 on the basis of whether the ovary is inferior or superior. More recently, Burtt, after prolonged study of the family, felt the need to alter this classification and proposed a division based on a rather unusual seedling character, the development of the plant from a single cotyledon (anisocotyly). Burtt's new division (1962) followed the geography of the family, the Old World species being anisocotylous, the New World species isocotylous. I have studied the chemistry of leaf and petal pigmentation in some 80 species, representing 40% of the genera, using material generously supplied by B. L. Burtt, and I have found four chemical characters (Table 4), the distribution of which clearly follows the new classification of the family by Burtt but cuts across the earlier system of Fritsch. Perhaps the most interesting of these is 3-deoxyanthocyanin, a class of pigment which occurs in almost all the isocotylous species but is lacking from every anisocotylous species so far examined. I was at first puzzled why 3-deoxyanthocyanins such as apigeninidin (Formula 8) and luteolinidin (Formula 9)

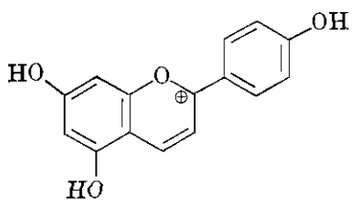
* Alkaloids in any case are such a heterogeneous group of secondary constituents that it is probably not wise to treat them as a single type of taxonomic character, but as several, depending on their particular biogenetic origin.

TABLE 4. Correlations between pigment chemistry and systematics in the Gesneriaceae

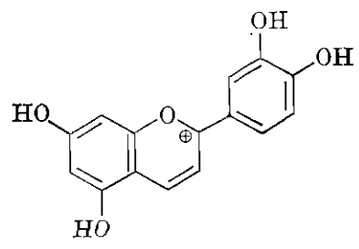
Pigment classes	Distribution in the subfamily	
	Gesnerioideae (New World, isocotylous)	Cyrtandroideae (Old World, anisocotylous)
Red deoxyanthocyanin	Present in 29 of 36 species	Absent from all of 37 species
Yellow carotenoid	Present in 5 of 36 species	Not detected
Yellow chalcone or aurone	Not detected	Present in 12 of 37 species
Yellow or red quinone	Not detected	Present in 8 of 37 species

Generic coverage = 40%. Taxonomy after Burtt (1962). Chemical data from Harborne (1966, 1967b).

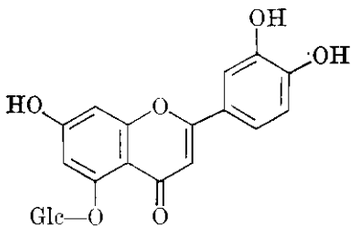
should only be present in New World species and then it occurred to me that the production of these orange–red pigments might be a reflection of natural selection in these mainly tropical plants for the red colour



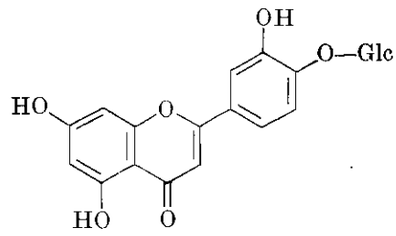
(8) Apigeninidin



(9) Luteolinidin



(10) Luteolin 5-glucoside



(11) Luteolin 4′-glucoside

favoured by the bird pollinators. Although bird pollination in the closely related Bignoniaceae is well recorded (see van der Pijl, 1961), I could at first find no account of it in the Gesneriaceae. However, I was able, by making enquiries, to obtain two independent confirmations

(Sutton, personal communication; Morley, 1966) that Jamaican gesnerads, such as *Gesneria* and *Achimenes*, are visited by the Bananaquit and other Humming Birds. By contrast, Old World species, such as *Streptocarpus*, must be insect-pollinated flowers and hence their blue and mauve petal, with colours produced by ordinary anthocyanins such as delphinidin and its derivatives.

In the Gesneriaceae, the reason why chemistry is related to geography is clearly a matter of natural selection for different pollination mechanisms. *Other examples of correlations between chemistry and geography* have come to light in recent years but in most cases it is not at all evident why such correlations should occur. For example, in the Eucryphiaceae, which consists of a single genus of five species, the two South American species, *E. glutinosa* and *E. cordifolia*, are very clearly distinguished (Table 5) by their leaf flavonoids from the three Austral-

TABLE 5. Distribution of flavonoids in the Eucryphiaceae

Flavonoid[a]	Identification	Distribution in *Eucryphia*				
		cordifolia	*glutinosa*[b]	*moorei*	*milliganii*	*lucida*
C	Caryatin	+	+	−	−	−
A 1	Az 3-galactoside	−	+	−	−	−
A 2	Az 3-diglucoside	+	+	−	−	−
A 3	Az 3-arabinosylgalactoside	+	−	−	−	−
A 4–6	Azaleatin (Az)	+	+	−	−	−
E	Unidentified flavonone (?)	+	+	−	−	−
Q 1	Quercetin 3-galactoside	+	+	−	+	−
Q 2	Quercetin 3-rhamnoside	+	+	−	−	−
Q 3	Quercetin 3-diglycoside	+	+	−	−	−
Q 4	Quercetin 3-triglycoside	−	−	+	−	−
D	Dihydroquercetin 3-glycoside(s)	−	−	−	+	−
L	Kaempferol 3,7-dimethyl ether	−	−	−	−	+
B	Cordifolia constituent	+	−	−	−	−
U	Unidentified flavan	+	+	−	−	−

[a] Colours in u.v. light without and with NH_3: C, blue (green); A 1–3, blue fluorescence (yellow); A 4–6, E, yellow fluorescence (yellow fl.); Q 1–3, dark ochre (dull green); L, dark ochre (dull yellow); B, colourless (green); U, absorbing (absorbing).

[b] The hybrid *E. nymansensis* is similar to *glutinosa* and, in addition, has component A 3.

(From Bate-Smith, Davenport and Harborne (1967). Reproduced by kind permission of Pergamon Press)

asian species, *E. moorei*, *E. lucida* and *E. milliganii*. All five are very similar morphologically—all have white flower colour—and it is difficult to see why the flavonoid pattern should be so distinctive. Other examples of correlations between flavonoids and geography occurring in the Rosaceae (*Pyrus*) (Williams, 1964) and in the Iridaceae (*Crocus*) (Bate-Smith and Swain, 1965) are equally difficult to explain unless the

flavonoids have some as yet unidentified protective function which is related to the habitat of the plant.

Chemosystematics of the Umbelliferae

In conclusion, I wish to say a few words about the chemosystematics of the Umbelliferae. A tribe of this family is being intensively studied at Liverpool from every viewpoint, as part of a multivariate approach to the numerical taxonomy of higher plants. One of the main features of the chemical programme is that we are concerned with studying every possible class of secondary constituent that varies in distribution and structure within the group. So far, we have concentrated on flavonoids and other water-soluble constituents but we are developing screening programmes for the lipid soluble constituents, especially furocoumarins, acetylenic compounds and terpenoids and for the seed proteins. The surveys of the family are by no means complete, but some interesting results have already come to light and I will mention briefly the relationship of the flavonoid pattern in leaf and flower to the systematics.

In choosing the spiny-fruited umbellifers to work on, we immediately run into the difficulty that there are two contrasting classifications from which to choose. While Bentham and Hooker place all these plants into a single tribe, the Caucalineae, Melchior in the latest revision (1964) of Engler's syllabus divides them into two quite separate groups, in part of the tribe Scandiceae and in part the Dauceae. That we are justified in considering the group as one follows from an extensive study of the distribution of the flavonoids and coumarins in umbellifer leaves. The Caucalineae *sensu* Bentham and Hooker stand out as a distinct chemical group. While the general rule in all the other tribes is the presence of kaempferol and/or quercetin as the major leaf flavonoid, in this tribe the flavonols are almost completely replaced by flavone and luteolin is a ubiquitous constituent. Again, furocoumarins, widely distributed in seed and root of most umbellifer tribes, are relatively rare in the Caucalineae. Chemistry thus tends to support the homogeneity of the Caucalineae and does not fit in with the Engler treatment.

Flavonoid chemistry has also been of value in illuminating relationships at the generic level in the Umbelliferae. For example, a very rare luteolin glycoside, the 5-glucoside galuteolin (formula (10)), has been detected in leaves of just two genera *Chaetosciadium* and *Torilis* of the Caucalineae (Harborne, 1967a). Superficially, the genera appear different, since fruit morphology—normally a key character—separates them; *Chaetoscia dium* has hairs on the fruit surface instead of spines. However, the chemical relationship is clearly significant since there are some good biological reasons for putting them close to each other. In particular,

they are the only genera in the group with a chromosome number of 12 and are very similar in their fruit anatomy. Again, a second rare luteolin glycoside, the 4′-glucoside (formula (11) has been found in the seed of the monotypic *Turgenia latifolia*, a plant which was once united with *Caucalis* but which our numerical analysis confirms as a separate genus. Again, there is a correlation with cytology, as *Turgenia* stands apart from the other taxa in having a chromosome number of 16.

Very recently, Dr. R. K. Crowden, who is working in my laboratory, has examined protein patterns in the seeds of members of the Caucalineae by acrylamide gel electrophoresis using the slab technique, and it is most satisfying to find that his studies confirm in most details the flavonoid results. Thus, in the chemical classification of this tribe of umbellifers the macromolecular and low molecular weight data support each other in indicating where the generic limits should be drawn.

REFERENCES

Alston, R. E. and Turner, B. L. (1963). "Biochemical Systematics" Prentice Hall, New Jersey.

Bate-Smith, E. C. and Swain, T. (1965). Recent developments in the chemotaxonomy of flavonoid compounds. *Lloydia* **28**, 313–31.

Bate-Smith, E. C., Davenport, S. M. and Harborne, J. B. (1967). A correlation between chemistry and plant geography in the genus *Eucryphia*. *Phytochemistry* **6**, 1407–14.

Burtt, B. L. (1962). The Gesneriaceae of the Old World. Keys to the tribes and genera. *Notes R. bot. Gdn Edinb.* **24**, 205–20.

Erdtman, H. (1956). Organic chemistry and conifer taxonomy. *In* Todd, A. R. (ed.) "Perspectives in Organic chemistry" 453–94. Interscience, New York.

Fritsch, K. (1893–4). Gesneriaceae. *In* Engler and Prantl (eds.) "Die Natürlichen Pflanzenfamilien" **4** (3B), 133–85.

Harborne, J. B. (1960). Two new naturally occurring anthocyanidins. *Chemy Ind.* 229–30.

Harborne, J. B. (1966). 3-Desoxyanthocyanins and their systematic distribution in ferns and gesnerads. *Phytochemistry* **5**, 589–600.

Harborne, J. B. (1967a). Luteolin 5-glucoside and its occurrence in the Umbelliferae. *Phytochemistry* **6**, 1569–73.

Harborne, J. B. (1967b). Flavonoid patterns in the Bignoniaceae and Gesneriaceae. *Phytochemistry* **6**, 1643–57.

Hegnauer, R. (1966). Comparative phytochemistry of the alkaloids. *In* Swain, T. (ed.) "Comparative Phytochemistry" 211–30. Academic Press, New York.

Hörhammer, L., Wagner, H., Bittner, G. and Graf, E. (1965). Separation of commercial *Aloe* samples by thin layer chromatography. *Dt. Apoth.Ztg* **105**, 827–30.

Marks, G. E., McKee, R. K. and Harborne, J. B. (1965). Double chromosome reduction in a tetraploid *Solanum*. *Nature, Lond.* **208**, 359–61.

Melchior, H. (1964). *In* Engler and Prantl (eds.) "Syllabus der Pflanzenfamilien" Vol. 2. Springer-Verlag, Basel.

Mirov, N. T., Zavarin, E. and Snajberk, K. (1966). Chemical composition of Mediterranean pines. *Phytochemistry* **5**, 97–102.

Morley, B. D. (1966). *Columnea* and aspects of its evolution. *J. Ass. Sci. Teach., Jamaica* **2**, 13–14.

Pijl, L. van der (1961). Ecological aspects of flower evolution. Part II. *Evolution, Lancaster, Pa.* **15**, 44–59.

Regnier, F. E., Waller, G. R. and Eisenbrau, E. J. (1967). Composition of essential oils of three *Nepeta* species. *Phytochemistry* **6**, 1281–9.

Rowlands, D. G. and Corner, J. J. (1963). Biochemical differences as varietal characteristics for peas, beans and spinach. *Hort. Res.* **3**, 1–12.

Swain, T. (ed.) (1963). "Chemical Plant Taxonomy." Academic Press, New York.

Swain, T. (ed.) (1966). "Comparative Phytochemistry." Academic Press, New York.

Wendelbo, P. (1961). *Primula* subgenus *Sphondylia* with a review of the subdivisions of the genus. *Arbok f. Univ. Bergen, Mat-Natur v. ser.* No. 11.

Williams, A. H. (1964). Flavone Glycosides of *Pyrus* species. *Chemy Ind.* 1318.

17. Chemistry and Taxonomy of Plants

E. C. Bate-Smith

A.R.C. Institute of Animal Physiology, Babraham, Cambridgeshire, England

ABSTRACT

Attention is drawn to recent publications by R. Hegnauer which cover this subject very fully.

From the historical point of view, chemistry has been used in plant taxonomy to a greater extent than is usually appreciated. This is especially true of the classification of the fungi, particularly as regards odour. Examples are given of the systematic distribution of specific odorous substances.

The interpretation of chemical data is helped very much by an understanding of the function of particular classes of substances in plants, especially in cases where the function is concerned with growth and differentiation. As an example, the recent discovery of the part played by the mono- and dihydroxyphenolic constituents in the control of auxin action in plants is discussed, and related to the systematic distribution of phenolic compounds.

This subject has recently been dealt with so adequately by Hegnauer (1967) that I propose merely to summarize his treatment and elaborate two or three of the points he raises.

Hegnauer has the advantage of being a pharmacognosist, that is, he has an equal interest in the chemistry of plants and in their identification, which of course requires a knowledge of taxonomy. He is the author of the first ever treatise of chemotaxonomy, four volumes of which have already been published (1962–66). Needless to say he is "sold" on chemotaxonomy, otherwise he would never have embarked on such an exacting task.

In his article (1967), which deals with both the possibilities and the limitations of chemical characters in plant taxonomy, Hegnauer attributes to de Candolle (1804) the first expression of a belief in the

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 193–199.

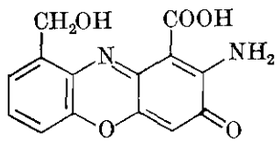
future value of chemical characteristics to plant taxonomy, reaffirmed in a quotation by Greshoff in 1890. From among many examples that might now be quoted in justification of their belief, Heguauer cites the chemical evidence in favour of revised positions proposed for three dicotyledouous families, the Callitrichaceae, the Cornaceae and the Hippuridaceae, by a number of systematic botanists. Under the heading of chemical characters as aids in delimitations he discusses the monocotyledonous families Liliaceae and Amaryllidaceae, and then gives a number of examples of their help in unambiguous identification of species. That he is equally well aware, however, of the need for caution is indicated by the space he devotes to the factors which limit the taxonomic value of chemical characters: to parallel evolutionary trends; to evolutionary convergence and divergence; and especially to the incorrect identification of plant samples!

A HISTORICAL PRECEDENT

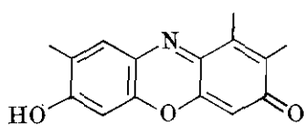
The first of Hegnauer's points I propose to take up is the historical one. In the world of botany, chemical characters have always been freely used by the taxonomist, even though they may not have been so regarded by him. Any specific distinction based on colour is a chemical distinction; even more so is one based on taste or odour. In the more familiar taxonomy of the higher plants these characters are not often regarded as critical, but in the case of the fungi they form an essential element in classification and differentiation. I am going to discuss the case of the fungi in some detail because this makes my point that chemotaxonomy does already exist, and in the narrowest and most conservative meaning of the term taxonomy.

As a matter of fact, I cut my own taxonomic teeth on the fungi, many years ago, and the first thing I learnt, from Cooke's textbook (1871), was that the agarics are *classified* according to the colour of their gills and spores. Going a little further, the colours of the pileus and stem are specific characters, and very often the odour of the plant is specified. At this stage, of course, it would be wrong to talk about chemotaxonomy, but these are the raw materials out of which chemotaxonomy can be—and is being—constructed. For instance, the distribution of the naphthaquinone and anthraqninone pigments in the Fusariae, the Boletaceae, the Polyporaceae and the Cortinariae has been extensively studied if only because of the obtrusiveness of these pigments and the consequent attraction of chemists to work on them. As might be expected, similar pigments to those present in fungi are found in lichens. Litmus, for instance, is chemically related to cinnabarin, the

striking orange-red pigment of *Trametes cinnabarina* and *Polystictus cinnabarinus* (Seshadri, 1966).



(1) Cinnabarin



(2) Chromophore of litmus

Commenting on the usefulness of chemistry in lichen systematics, Dahl (1950) remarked "most of the more important sub-divisions of Parmelia are confirmed and decisive arguments may be found in many dubitable cases". He cites this as a "beautiful" example. Characteristically, the only recorded comment on his paper (H. des Abbayes) was "il faut être très prudent dans la généralisation de cette méthode".

At least as much as colour, odour is employed in the taxonomy of the fungi as a distinctive character. From the broadest point of view it is possible to speak of a "fungal" odour—or of a musty odour especially characteristic of Actinomycetes. These are not just subjective characters, they are truly chemical characters, since the substances responsible for the odours are specific—in a number of instances, already known—chemical substances. One such substance, the odour of which is easily recognizable as the characteristic odour of mushroom, has been isolated from *Armillaria matsutake* and identified as vinylamylcarbinol (Murahashi, 1936, 1938):

$CH_3—(CH_2)_4—CH(OH)—CH=CH_2$ "Matsutake alcohol" $CH_3—CH_2—CH=CH—CH_2—CH_2OH$ "Leaf alcohol"

The vinylethyl compound has an odour very similar to that of *Scleroderma vulgare*. It is interesting to note that the smell of the leaves of higher plants is often due to the presence of hexenol and hexenal, known in German as *Blätteralkohol* and *Blätteraldehyd*. These distinctive constituents of the fungi and the angiosperms are truly chemotaxonomic characters referable to taxa at a very high level. It is, however, at the generic and specific level that odour characters are most employed by mycologists. In "Les Champignons d'Europe", Roger Heim (1957), the Director of the Muséum National d'Histoire Naturelle writes: "certain naturalists, especially French ones such as Quélet, have successfully employed olfactory characters in the diagnosis of the higher fungi. Others, e.g. Gilbert, have endeavoured to use these properties in the systematisation of species, or to define the odours of fungi in chemical terms." I have myself, during the last few years, in collaboration with an experimental psychologist and a biochemist, been struggling with the problem of the interpretation of odour in chemical terms, and I am

optimistic that before very long we shall be able to do this, and that "olfactory characters" will be incorporated into chemical taxonomy. Before this can be done, however, it will be necessary to devise a biotaxonomy of the chemical constituents, that is the replacement of the traditional *chemical* classification of the constituents of living organisms by one which takes into account the manner of their formation and the mechanism of their physiological action. So far as the first of these is concerned, the "premier pas" was taken by Charles Mentzer (in Mentzer and Fatianoff, 1964; Mentzer, 1966) in his "Actualités de Phytochimie fondamentale". Of the sensory properties of matter, odour and taste are the most intimately concerned with chemistry. The specific olfactory characters of plants, which are as easily observable as colour and form, are due to the presence of specific chemical substances in those plants, and it is very much worth while, therefore, making an attempt to identify the substances responsible for such specific olfactory characters, and to relate their presence to processes which are peculiar to the organisms that fabricate them. We can, for instance, try to discover why hexenols and hexenals are elaborated by angiospermous plants and vinylcarbinols by fungi, and relate the biosynthetic processes so disclosed to the particular life-forms of the two kinds of plants. There is, in this connection, a tendency to shift the onus of responsibility on to the enzymologist to identify the enzyme systems operating in the formation of specific constituents; for the enzymologist to shift the responsibility on to the molecular biologist to designate the RNA system operating in the cytoplasm, the DNA system operating in the nucleus, and the genetic code which determines the kind of plant the plant is. The chemical taxonomist can, however, make use of the ample information at present available in the array of substances actually encountered in plants, so long as the information is appropriately recoded.

In the case of odour, coding has at present to be done more in terms of odour qualities than of specific chemical substances; although there are, of course, cases where odours can be attributed to specific substances. Examples of this are clove, attributable to eugenol, vanilla, coumarin and violet, the last attributable to a small group of substances of which α- and β-ionone are the most familiar. Chemical identification is the soundest—in fact the only unexceptionable—basis for the employment of odour in chemotaxonomic exercises. Many of the epithets applied to the olfactory characters of fungi (see Heim, *loc. cit.*, for a detailed discussion of such epithets) are of this kind, but few of the odorous constituents have as yet been rigorously identified. There are many instances, however, where the odour is restricted in its incidence to particular groups of fungi which are regarded as nearly related. One such instance, quoted by Heim, is that of three neighbouring genera,

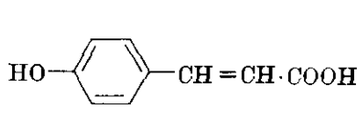
Cortinarius, *Hebeloma* and *Inocybe*, where the characters of odour of radish, fruit and piperidine are distributed, approximately, so that each is present only in two of the three genera.

Since by their nature odorous substances are volatile, and since many are released at particular times or in particular circumstances, it may be that it is a complex precursor in the plant rather than the odour itself which should be looked for to provide taxonomic information. A case in point is the odour of eugenol which becomes apparent when the roots of *Geum urbanum* are bruised. This arises from a glycoside of eugenol, gein, present in the roots and not as a constituent of an essential oil, the form in which eugenol usually occurs, and as such probably having quite a different biosynthetic origin.

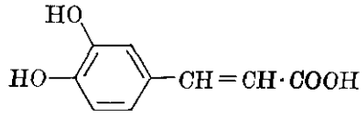
A TOPICAL EXAMPLE

The usability of chemical taxonomic information will be greatly increased as more comes to be known about the function in the plant of the so-called secondary constituents. Until quite recently, phytochemistry has been little more than a catalogue of what substances occur in what plants. The principles, if any, guiding their systematic distribution have been completely hidden. As particular categories of substances become linked with particular aspects of anatomical structure or physiological function, those categories, and individual members of them, will automatically fall into place as evidence upon which systematic relationships can be adduced. A striking example of this is actually happening at the present moment.

Two years ago (Bate-Smith, 1965) it was known (Tomaszewsky, 1963) that *p*-coumaric and caffeic acids:

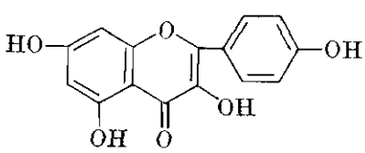


(3) *p*-Coumaric acid

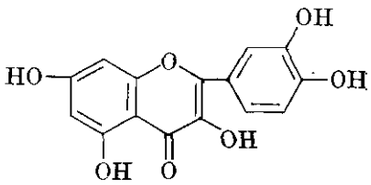


(4) Caffeic acid

had opposing effects on the activity of indoleacetic acid oxidase, as also had the flavonols kaempferol and quercetin (Smith *et al.*, 1961; Furuya *et al.*, 1962):



(5) Kaempferol



(6) Quercetin

Since that date, Tomaszewsky and Thimann (1966) have produced detailed evidence of the mechanism of the effect in the case of the phenolic acids, concluding that their experiments "point to a major role for phenolase in controlling hormone balance, since the introduction into a phenolic molecule of a second, adjacent hydroxyl group changes the action *from auxin-destroying to auxin-preserving*. Thus the phenol oxidizing enzymes must act as general growth controllers".

The recent results of Jaffe and Galston (1967) are even more exciting. Following indications previously reported by Tronchet (1961), they have shown that pea tendrils contain high concentrations of glycosides of kaempferol and quercetin in which the sugar residue has a highly specific and unusual configuration, viz. triglucosyl-*p*-coumarate. When the tendrils are stimulated, and during their coiling, the concentration of the quercetin glycoside falls to as little as 30% of its original value, that of the kaempferol glycoside remaining almost unchanged. Good reasons are presented for considering that the quercetin glycoside is actively concerned in the movement of the tendril by way of an ATP *enzyme system*.

From the systematic point of view, the significance of this observation lies in the fact that the presence or absence of quercetin and kaempferol is highly correlated with the systematic position of the plants in which they occur, or do not occur. When they are absent their place is taken (or appears to be taken) by other flavonoid compounds differing in such respects as the absence of a hydroxyl group in the 3-position (flavones); reduction of the central ring (flavanones); absence of a 5-hydroxyl group (restricted to the Leguminosae, Anacardiaceae and Compositae); presence of additional hydroxyl groups; or, of especial significance, C-glycosylation in the 6- or 8-position. Like the flavonols, these substances all occur in the plant as glycosides and they also all occur, like kaempferol and quercetin, in mono- and dihydroxy substituted form. Their configurational modifications may therefore be seen as devices for restraining some, while retaining most, of the functions which the flavonols normally fulfil. Different plants employ different devices, according to their kind.

REFERENCES

Bate-Smith, E. C. (1965). Recent progress in the chemical taxonomy of some phenolic constituents of plants. *Bull. Soc. bot. Fr., Mémoire sur Chimiotaxonomie*, 16–28.

de Candolle, A. P. (1804). Essai sur les propriétés médicales des plantes, *comparées avec leurs formes extérieures et leurs classification naturelle*. Paris.

Cooke, M. C. (1871). "Handbook of British Fungi." Macmillan, London and New York.

Dahl, E. (1950). Use of lichen chemistry in lichen systematics. *Proc. 7th Int. bot. Congr. Stockholm*, 810.

Furuya, M., Galston, A. W. and Stowe, B. B. (1962). Isolation from peas of co-factors and inhibitors of indolyl-3-acetic acid oxidase. *Nature, Lond.* **193**, 456.

Greshoff, M. (1890). Koninkl. Natuurk. Veren. in Ned.-Indië (voordrachten No. 1) "Planten en plantenstoffen". Ernst, Batavia and Noordvijk.

Heim, R. (1957). "Les Champignons d'Europe" Vol. 1. N. Boubée, Paris.

Hegnauer, R. (1962–66). "Chemotaxonomie der Pflanzen" Vols 1–4. Birkhäuser Verlag, Basel and Stuttgart.

Hegnauer, R. (1967). Chemical characters in plant taxonomy: some possibilities and limitations. *Pure appl. Chem.* **14**, 173–87.

Jaffe, M. J. and Galston, A. W. (1967). Physiological studies on pea tendrils. IV. Flavonoids and contact coiling. *Pl. Physiol., Lancaster* **42**, 848–50.

Mentzer, C. and Fatianoff, O. (1964). "Actualités de Phytochimie Fondamentale." Masson, Paris.

Mentzer, C. (1966). "Actualités de Phytochimie Fondamentale" series II. Masson, Paris.

Murahashi, S. (1936). The odour of matsutake. I. *Sci. Pap. Inst. phys. chem. Res., Tokyo* **30**, 263–71.

Murahashi, S. (1936). The odour of matsutake. II. *Sci. Pap. Inst. phys. chem. Res., Tokyo* **34**, 155–72.

Seshadri, T. R. (1966). Verbal communication to the Ceylon Association for the Advancement of Science.

Smith, D. H., Mumford, F. E. and Castle, J. E. (1961). IAA oxidase inhibitor from etiolated pea epicotyls. *Pl. Physiol., Lancaster* **36**, xlvii.

Tomaszewsky, M. (1963). *Colloques int. Cent. natn. Rech. scient.* **123**, 335–51.

Tomaszewsky, M. and Thimann, K. V. (1966). Interactions of phenolic acids, metallic ions and chelating agents on auxin-induced growth. *Pl. Physiol., Lancaster* **41**, 1443–51.

Tronchet, J. (1961). Sur la presence de flavonoides dans les tissues corticaux des vrilles et des tiges volatiles, et leur intervention possible dans le circumnutation. *86th Congrès Soc. Savantes, Montpelier*, 493–5.

18. Biochemical and Immunological Studies on the Evolution of a Metabolic Pathway in Bacteria

R. Y. Stanier

Department of Bacteriology and Immunology, University of California, Berkeley, California, U.S.A.

ABSTRACT

The commonest metabolic pathway for the oxidative dissimilation of aromatic compounds, the β-ketoadipate pathway, is widely distributed through groups of aerobic bacteria which, on other phenotypic grounds, do not appear to be closely related. A comparative analysis of the regulatory mechanisms governing synthesis of the inducible enzymes operative in this pathway has shown that *Pseudomonas* spp. all share a particular system of regulation which is different from the system of regulation employed by *Moraxella* (*Acinetobacter*) to govern the enzymes of this pathway. It seems likely that the comparative analysis of control systems may provide a new and valuable parameter for the study of biochemical evolution. By the use of antisera prepared against specific enzymes of the pathway, it could also be shown that antigenic homology with respect to some enzymes of the pathway is broadly correlated with homology of the regulatory machinery.

INTRODUCTION

The very heterogeneous collection of micro-organisms termed "bacteria" offers tantalizing but formidable challenges to the taxonomist and the evolutionist alike. Although there is no fossil evidence on which to base a history of this group, the extreme antiquity of the bacteria is suggested by their distinctive and relatively undifferentiated internal cellular organization, termed *procaryotic*. The procaryotic type of cell occurs only in bacteria and in blue-green algae; all other contemporary biological groups have cells of the more familiar, eucaryotic type (Stanier and van Niel, 1962; Stanier, 1964). Another trait indicative of

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 201–225.

antiquity (as well as of major internal diversity) is the remarkably wide span of DNA base composition, ranging for the bacteria as a whole from approximately 30 to approximately 70 moles % of guanine and cytosine (Hill, 1966).

From the earliest days of bacterial taxonomy, it has been recognized that structural characters *alone* are insufficient to permit a satisfactory subdivision of the bacteria (Cohn, 1872). Nor has the advent of electron microscopy caused any substantial change in this situation. The insights into cell structure obtained by the use of the electron microscope proved indispensable for the definition of the general properties of procaryotic cells, but have not greatly increased the number of structural characters available for the internal subdivision of bacteria. For this reason, bacteriologists have been forced to use functional characters far more extensively than has hitherto been customary in other branches of biology to describe phenotypes and to achieve workable taxonomic subdivisions.

The task of selecting functional attributes appropriate for taxonomic purposes is a delicate one, which must be guided by the insights obtainable from biochemistry and genetics. It was not always wisely performed in the early days of bacterial taxonomy. For example, the ability to ferment the disaccharide lactose has been used as an important generic trait among the coliform bacteria, but is completely unsuitable for this purpose, since it is the expression of a single structural gene, governing production of the enzyme β-galactosidase. The kinds of functional attributes which are appropriate as major taxonomic markers among bacteria are distinctive *metabolic unit processes*; namely, complex functions that are determined by integrated, multienzyme systems, subject to multigene control. As examples, one could cite the possession of a specialized fermentative pathway, the capacity to perform photosynthesis, the capacity for chemo-autotrophy, the nature of the respiratory electron transport system.

In recent years, molecular biology has provided a completely unexpected and extremely valuable control on the homogeneity of bacterial taxa previously recognized in terms of structural and functional attributes. I have already alluded to the exceptionally wide span of DNA base composition characteristic of the bacteria as a whole. The very extensive body of analytical data concerning DNA base ratios of bacteria that is now available (Hill, 1966) has revealed that many genera or generic clusters earlier recognized by bacterial taxonomists show a rather narrow and characteristic internal span of DNA base composition. This is true, for example, of the coliform bacteria, the actinomycetes, the sporeformers, the fruiting myxobacteria. Since DNA base composition appears to be a stable character, as evidenced

by its essential uniformity within a species (Mandel, 1966), we can begin dimly to perceive amongst the bacteria certain genera, or occasionally generic clusters, which may well represent evolutionary units as well as taxonomic ones.

THE PROBLEM OF HIERARCHICAL CLASSIFICATION AMONG BACTERIA

The taxonomists who believe that the construction of suprageneric hierarchies among bacteria is a useful operation have always faced a dilemma: what relative weight should be assigned to structural and to functional attributes in the creation of families, orders and classes? The dilemma arises from the fact that the linkage between major structural and functional properties among bacteria is loose and irregular; not uncommonly, a distinctive functional attribute occurs in two or more groups that have markedly different structural attributes. For example, the capacity for dissimilatory sulphate reduction (the ability to perform an anaerobic mode of respiration with sulphate as terminal electron acceptor) characterizes two genera of anaerobic bacteria, *Desulfovibrio* and *Desulfotomaculum*, which differ from one another markedly in cell structure, and each of which shows structural resemblances to other bacterial groups which are not sulphate-reducers (Postgate, 1965). For the creation of bacterial families, it is accordingly necessary to decide whether these two genera should be united into a single family based on the rare and distinctive attribute of sulphate reduction, or placed in separate families, defined in terms of structural properties. One can argue interminably and fruitlessly about such taxonomic pseudo-problems. This explains why hierarchical schemes of bacterial classification are so unstable. The genus, defined on a combination of structural and physiological properties, and characterized by a relatively limited span of DNA base composition, at present appears to be the highest bacterial taxonomic unit that is definable in a logically consistent and useful fashion. It should be emphasized, however, that what the bacteriologist recognizes as a genus is almost certainly incommensurate with the taxon bearing the same designation in a more recently evolved biological group, such as the vertebrates or the spermatophytes. This is a point to which I shall return at the end of this paper.

The pseudo-problem of bacterial hierarchies does conceal a real and interesting problem concerning physiological evolution. Throughout the bacteria, so diverse in their functional attributes, rare and distinctive metabolic unit-processes like dissimilatory sulphate reduction crop up sporadically in groups which do not appear closely related to one

another in terms of other criteria. Two different evolutionary interpretations of this fact are possible. Conceivably, each of these metabolic unit-processes could have been a unique invention of early biochemical evolution, still preserved in two or more lines which have undergone marked divergence in other respects. Alternatively, a given metabolic unit-process, which appears homologous on gross biochemical examination, might have arisen independently, out of totally different genetic backgrounds, in two or more biological lines. Is it possible by more refined analysis to achieve the vital discrimination between analogy and homology with respect to a widely distributed bacterial metabolic unit-process? That is the question which I propose to examine here.

THE OXIDATION OF AROMATIC COMPOUNDS BY BACTERIA

Some ways of approaching this general problem are illustrated by the work from our laboratory on the bacterial utilization of aromatic compounds as substrates for respiratory metabolism. The capacity to use any aromatic compound as a source of carbon and energy implies possession of a complex and specialized piece of metabolic machinery, since it requires the operation of a large array of specific enzymes, acting in sequence, to convert the primary substrate into intermediates that lie on the universal pathways of cellular intermediary metabolism. These specific enzymes are all typically under inducible control. The ability to grow at the expense of aromatic compounds is a property that is widely dispersed among aerobic bacteria.

Our studies were initially undertaken on the members of the genus *Pseudomonas*: Gram-negative, polarly flagellated rod-shaped bacteria with a DNA base composition in the range of 60–70 moles % guanine and cytosine. Among these bacteria, the most commonly utilized aromatic compound is *p*-hydroxybenzoic acid, which can support the growth of many different species. However, we have found that this shared nutritional attribute does not even always reflect possession of a common biochemical machinery. Complete metabolic divergence can occur after the first step-reaction, as a result of differences among pseudomonads in the enzymatic mechanism of aromatic ring cleavage (Fig. 1). Furthermore, the divergent pathways for the metabolism of *p*-hydroxybenzoate have a very clear cut taxonomic distribution within the genus *Pseudomonas*, as shown in Table 1. Two related species, *P. acidovorans* and *P. testosteroni* (comprising a subgeneric cluster termed the acidovorans group), invariably employ the *meta*-cleavage pathway; all other *Pseudomonas* spp. the *ortho*-cleavage pathway. *P. testosteroni* and *P. acidovorans* likewise differ from other members of the genus with respect to the pathways that they employ for the

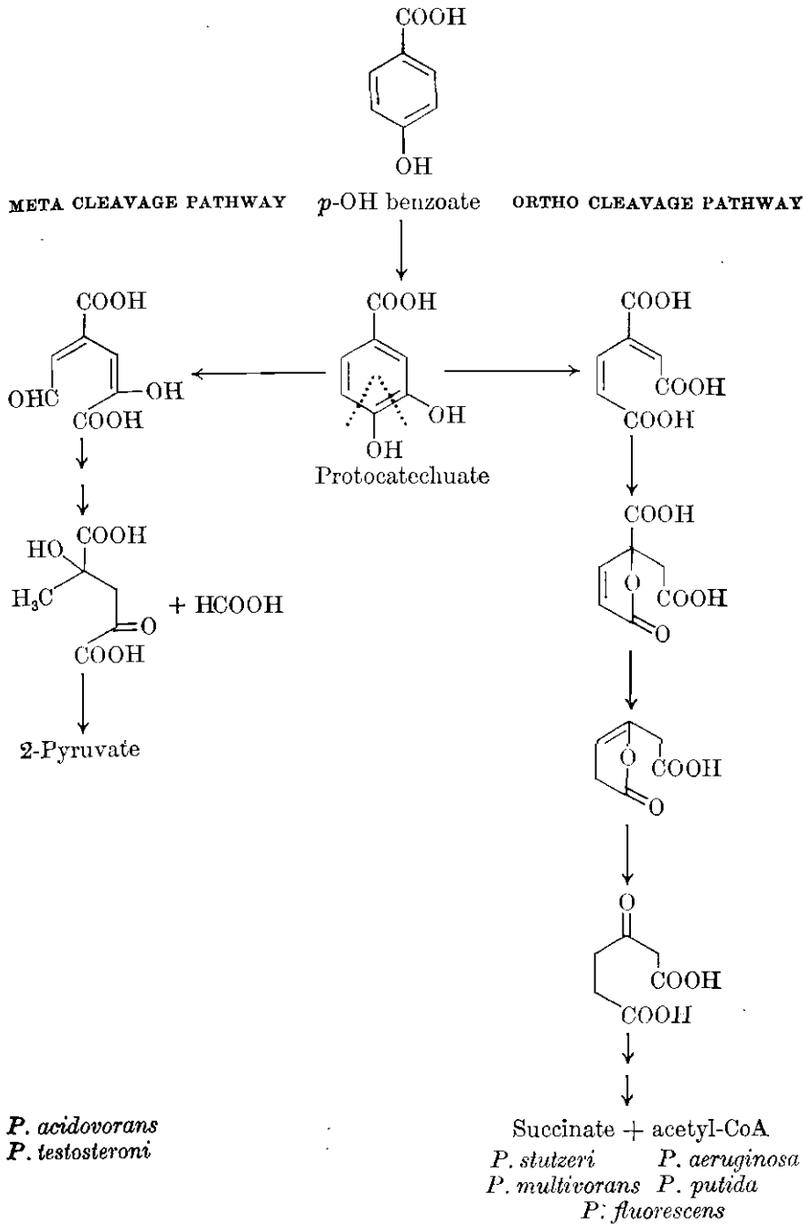


FIG. 1. The two pathways for the dissimilation of *p*-hydroxybenzoate in the genus *Pseudomonas*.

TABLE 1. Mechanisms for ring cleavage of protocatechuate among *Pseudomonas* spp.

Species-group	Species	Number of strains examined	Number of strains showing *ortho*-cleavage	Number of strains showing *meta*-cleavage
Fluorescent	*P. aeruginosa*	5	5	0
	P. fluorescens	62	62	0
	P. putida	34	34	0
	P. multivorans	8	8	0
	P. stutzeri	3	3	0
Pseudo-mallei	*P. pseudomallei*	5	5	0
	P. mallei	13	13	0
Acidovorans	*P. acidovorans*	15	0	15
	P. testosteroni	9	0	9

metabolism of benzoate and tryptophan, respectively, as shown in Figs. 2 and 3. Analysis on the simplest biochemical level accordingly reveals profound differences in enzymatic constitution between two groups in the genus *Pseudomonas*, differences which are hidden in terms of nutritionally defined phenotypes. The acidovorans group shares with other *Pseudomonas* species many additional nutritional characters that have not yet been analysed in terms of their underlying biochemical mechanisms, and it seems very likely that future work will bring to light additional biochemical dichotomies. If the total metabolic map of the acidovorans group is as singular as the presently known fragment of it suggests, it would seem reasonable to conclude that the acidovorans group represents an evolutionary line quite different from the main *Pseudomonas* line, for which separation on the generic level might well be desirable.

In other groups of aerobic bacteria where the mechanism of *p*-hydroxybenzoate oxidation has been studied biochemically (*Hydrogenomonas, Moraxella**, *Mycobacterium-Nocardia, Bacillus, Azotobacter*), the data (although often fragmentary) suggest that dissimilation proceeds through the *ortho*-cleavage pathway. This particular reaction-sequence is one component of a very complex set of convergent reaction sequences which are used by such bacteria for the oxidation of a variety of primary substrates, aromatic and hydroaromatic. In all cases, the six carbon atoms of the ring in the primary substrate eventually give rise to the six carbon atoms of β-ketoadipic acid (Fig. 4). β-Ketoadipic acid is the last specific intermediate, since it is cleaved to succinate and acetyl-CoA. For this reason, the whole metabolic network is termed the β-ketoadipate pathway. By no means every primary substrate shown

* See note on p. 225.

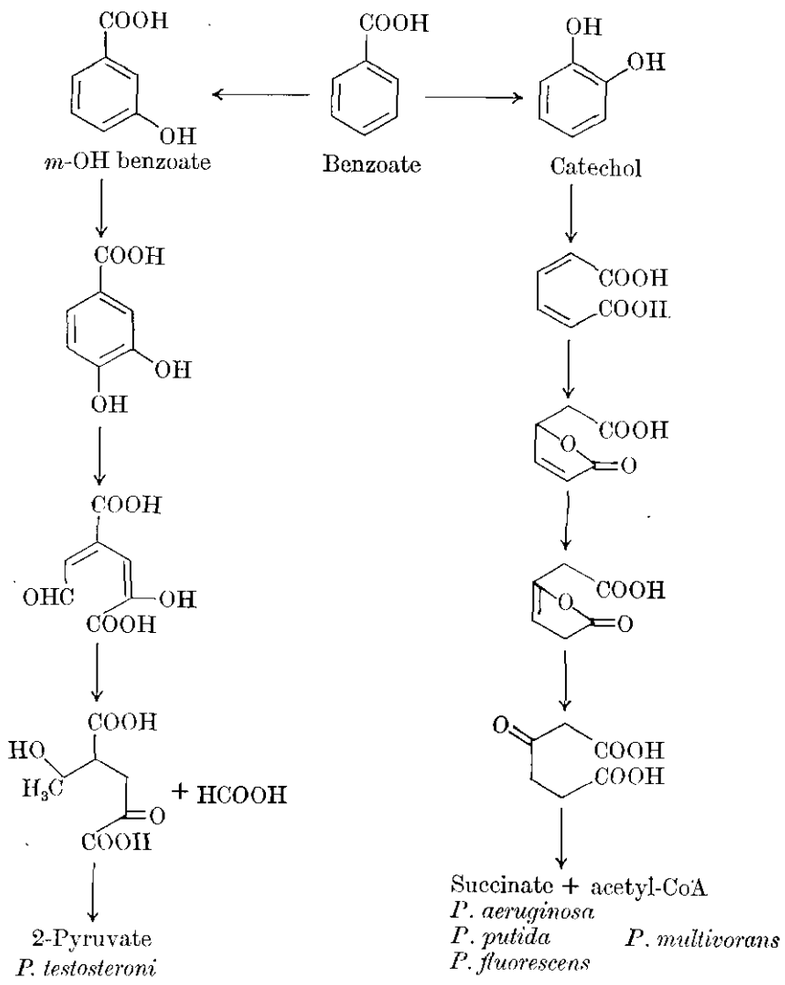


FIG. 2. Differences between the acidovorans and fluorescent groups with respect to the pathway for benzoate dissimilation.

in Fig. 4 can necessarily be attacked by any given bacterial strain or species that utilizes this pathway; but the heart of the metabolic network, comprising the parallel reaction-sequences leading from catechol and protocatechuate, together with the terminal sequence that gives rise to succinate and acetyl-CoA, are almost universally present.

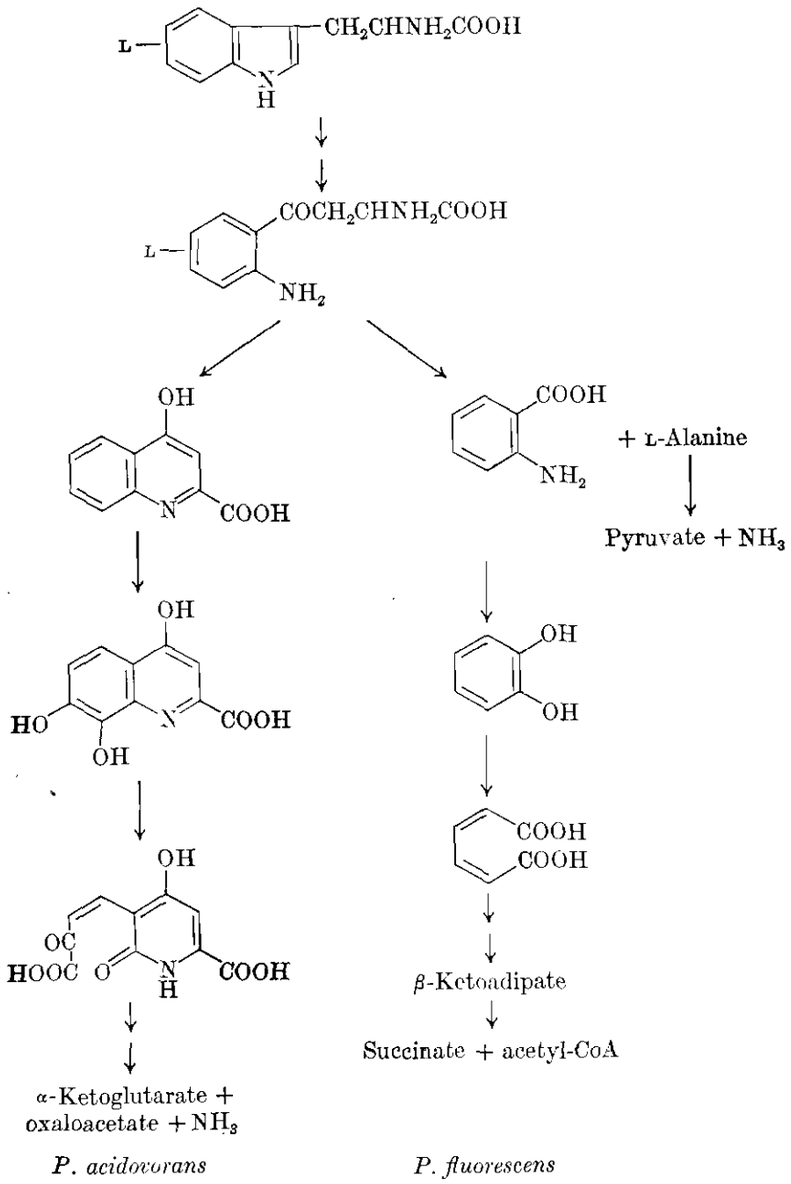


FIG. 3. Differences between the acidovorans and fluorescent groups with respect to the pathway for tryptophan dissimilation.

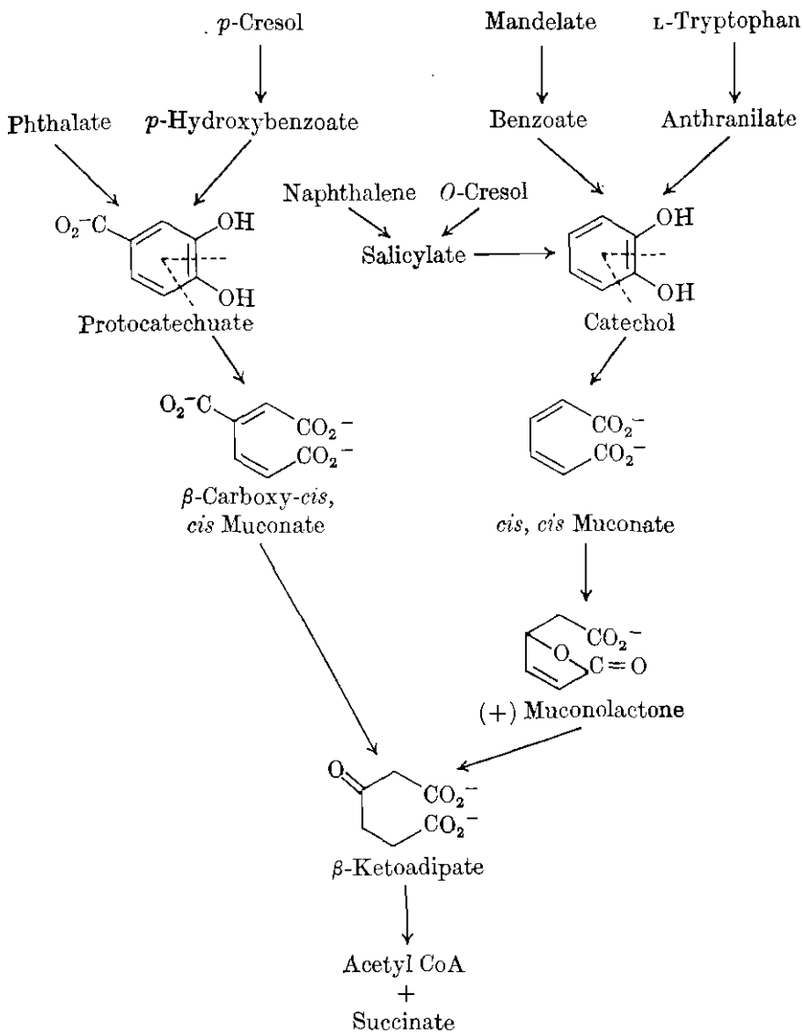


FIG. 4. A general view of the convergent reaction-sequences of the β-ketoadipate pathway.

THE SIGNIFICANCE OF CONTROL MECHANISMS

Does the wide taxonomic distribution of this somewhat complex piece of biochemical machinery mean that the β-ketoadipate pathway has had a single evolutionary origin among aerobic bacteria, or is its

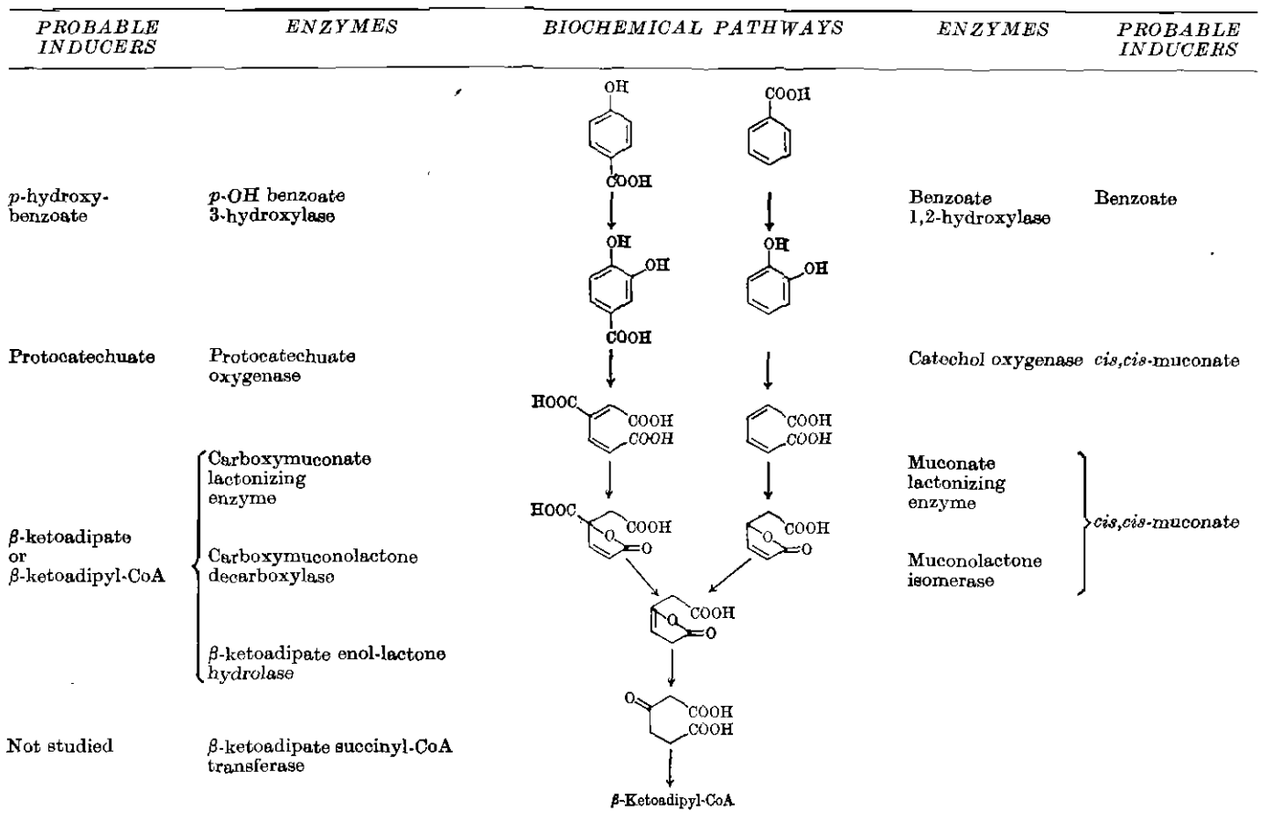


FIG. 5. Regulation of the β-ketoadipate pathway in *Pseudomonas putida*.

appearance in groups as diverse as *Pseudomonas*, *Hydrogenomonas*, *Moraxella*, *Bacillus* and *Mycobacterium* a consequence of parallel, convergent biochemical evolution? In attempting to analyse this question, we have found that the study of the control mechanisms governing synthesis of these inducible enzymes provides very useful information.

This approach to the problem was made possible as a result of the detailed analysis of the regulation of the β-ketoadipate pathway in a fluorescent pseudomonad, *P. putida*, by Ornston (1966b). The regulatory map which he established for this species is shown in detail in Fig. 5. One striking feature is the coordinate product-induction by β-ketoadipate (or possibly β-ketoadipyl-CoA) of a group of three enzymes that catalyse three sequential reactions immediately preceding the formation of β-ketoadipate itself. Two of these enzymes, carboxymuconate lactonizing enzyme and carboxymuconolactone decarboxylase, are specific to the protocatechuate branch of the pathway, while the third, enol-lactone hydrolase, mediates the first common step-reaction following metabolic convergence of the catechol and protocatechuate branches of the pathway. This curious regulatory linkage has an obvious physiological consequence: cells grown on any precursor of the catechol branch (e.g. benzoate) synthesize high levels of the last two enzymes specific to the protocatechuate branch, even though the two enzymes in question are non-functional under these conditions of growth. This feature of control can be very easily examined. We have found that it always occurs in other *Pseudomonas* species that employ the reactions of the β-ketoadipate pathway: specifically, the two other species of the fluorescent group, *P. aeruginosa* and *P. fluorescens*, *P. stutzeri* and *P. multivorans*. Some of these species can use adipate as a carbon and energy source, and induction of the last two enzymes of the protocatechuate branch is also elicited by this substrate, presumably as a result of the formation of β-ketoadipate (or β-ketoadipyl–CoA) as an intermediate in adipate metabolism (Table 2).

However, induction of carboxymuconate lactonizing enzyme and carboxymuconolactone decarboxylase by β-ketoadipate does not occur in representatives of other genera using the β-ketoadipate pathway that we have examined so far. In strains of *Hydrogenomonas* and *Moraxella*, these enzymes are not measurably induced as a result of growth on benzoate or adipate (Table 2). We have now explored in detail the regulation of the β-ketoadipate pathway in *Moraxella calcoacetica* (Cánovas and Stanier, 1967). This bacterium is a Gram-negative, non-flagellated rod-shaped organism quite similar in its nutrition and ecology to the aerobic pseudomonads. However, it has DNA of very different base composition, containing 40–45 moles percent of guanine and cytosine. The regulatory map for the β-ketoadipate pathway in *M. calco-*

TABLE 2. Influence of the growth substrate on synthesis of CMLE, CMD and ELH by several species of bacteria

Organisms	Enzyme activities	Growth substrates: Succinate	*p*-Hydroxybenzoate	Benzoate	Adipate
Pseudomonas fluorescens, biotypes A, B, D, E *Pseudomonas putida*, biotypes A, B	CMLE, CMD, ELH	−	+	+	not used
Pseudomonas aeruginosa *Pseudomonas fluorescens*, biotype C *Pseudomonas multivorans*	CMLE, CMD, ELH	−	+	+	+
Hydrogenomonas eutropha	CMLE, CMD	−	+	−	−
Moraxella calcoacetica	ELH	−	+	−	−

− = basal level; + = induced activity at least 20 times basal level. CMLE = carboxymuconate lactonizing enzyme; CMD = carboxymuconolactone decarboxylase; ELH = β-ketoadipate enol-lactone hydrolase.

acetica is shown in Fig. 6. The differences from the regulatory map for *P. putida* are numerous and profound. They can be summarized as follows:

1. Co-ordinate regulation is much more pronounced in *M. calcoacetica*. Enzymes of both the catechol and the protocatechuate branches are synthesized co-ordinately with later enzymes that mediate the terminal, common step-reactions of the pathway.

2. Rigorous physiological specificity of control over the enzymes of the two branches of the pathway is maintained by a biochemical device not found in *P. putida*: synthesis of two sets of isofunctional enzymes, under different inductive control, to mediate the terminal, common step-reactions. One set is synthesized co-ordinately with enzymes specific to the catechol branch of the pathway, the other set co-ordinately with enzymes specific to the protocatechuate branch.

3. There are major differences with respect to the intermediary metabolites that function as inducers, as well as the specific enzymes whose synthesis they control. β-Ketoadipate, which has a cardinal inductive function in pseudomonads, cannot serve as an inducer of *any* enzymes of the pathway in *M. calcoacetica*. As a consequence of its failure to serve as an inducer in this species, it cannot support growth and is, accordingly, a powerful agent for the selection of constitutive mutants. Protocatechuate, which in *P. putida* elicits synthesis of only one enzyme (protocatechuate oxygenase), acts in *M. calcoacetica* as the

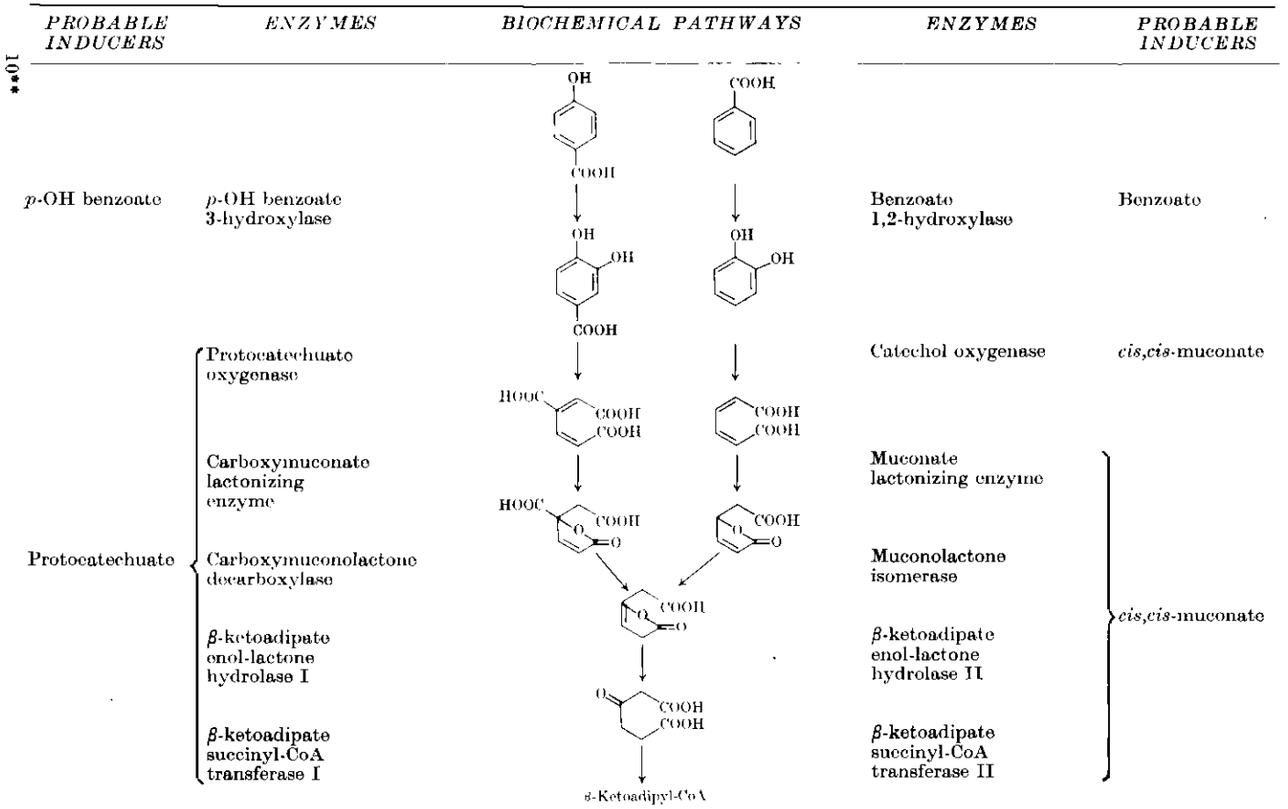


FIG. 6. Regulation of the β-ketoadipate pathway in *Moraxella calcoacetica*.

inducer of an extensive co-ordinate block, comprising no less than five enzymes.

Accordingly, a more refined biochemical analysis has revealed that the β-ketoadipate pathway, identical on the metabolic level in *Pseudomonas* spp. and *M. calcoacetica*, is markedly dissimilar with respect to its regulation, and even in part with respect to its enzymatic mediation. It is difficult to avoid the conclusion that this particular piece of biochemical machinery has had different evolutionary origins in the two groups. At first sight, this conclusion may appear surprising. Much of the speculation about biochemical evolution has been based on the assumption (implicit or explicit) that complex metabolic pathways do indeed serve as useful markers of evolutionary relatedness. It has never been sufficiently emphasized that strictly chemical considerations set severe constraints on the manner in which natural selection can solve any given biochemical problem. For example, if the goal of natural selection is to endow an organism with the ability to convert aromatic compounds into intermediates of the TCA cycle (essential for their utilization as sources of carbon and energy), the chemically permissible solutions may be very few in number. At the present time, the only known chemical means of performing this operation involves *ortho*-cleavage of the benzene ring, a device also widely adopted among micro-organisms. As already mentioned, however, in the acidovorans group of pseudomonads natural selection has discovered one additional solution of the problem (*meta* cleavage) which has not yet been achieved by the organic chemist. The *ortho*- and *meta*-cleavage pathways probably represent the only permissible solutions of this specific biochemical problem, and may well have been separately evolved in many different biological groups.

IMMUNOLOGICAL EVIDENCE FOR THE HOMOLOGY OF ENZYMES IN THE β-KETOADIPATE PATHWAY

Since a key control mechanism (co-ordinate product-induction of carboxymuconate lactonizing enzyme, carboxy-muconolactone decarboxylase and enol-lactone hydrolase by β-ketoadipate) is characteristic of all *Pseudomonas* spp. that employ the β-ketoadipate pathway, it is conceivable that *within* this genus, the enzymatic machinery of the β-ketoadipate pathway may have had a single evolutionary origin. This hypothesis is plausible, since the species in question share many other phenotypic traits, and are quite similar in DNA base composition. We have therefore attempted to obtain other kinds of evidence that bear on the question of infra-generic homology.

During his work on the pathway in *P. putida*, Ornston (1966a)

succeeded in crystallizing muconate lactonizing enzyme and muconolactone isomerase, enzymes which catalyse successive step-reactions in the catechol branch, and are co-ordinately induced by *cis,cis*-muconate, substrate for the former enzyme. The availability of these two enzymes in a state of purity has permitted us to prepare highly specific antisera against each of them. On gel diffusion plates, each antiserum yields a single, sharp precipitin line both with the homologous pure enzyme and with a crude, cell-free extract of benzoate-grown cells of the strain from which the immunizing antigen was isolated (Fig. 7). Extracts of this strain grown on succinate or on *p*-hydroxybenzoate do not contain

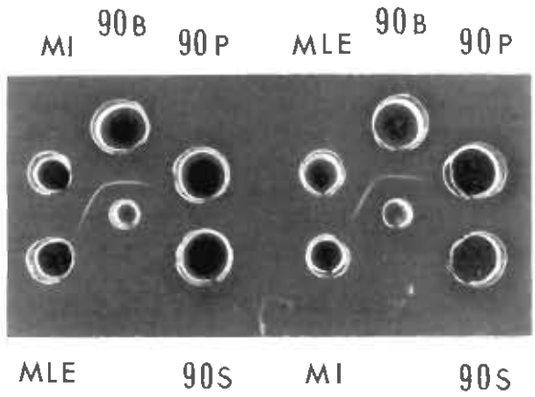


FIG. 7. The specificity of precipitin reactions with antisera directed against crystalline muconate lactonizing enzyme and crystalline muconolactone isomerase, prepared from *P. putida* strain 90. Left centre well: anti-muconolactone isomerase; right centre well: anti-muconate lactonizing enzyme. Contents of other wells: MLE, crystalline muconate lactonizing enzyme (1 unit); MI, crystalline muconolactone isomerase (5 units); 90B, crude extract of benzoate-grown *P. putida* 90; 90S, crude extract of succinate-grown *P. putida* 90; 90P, crude extract of *p*-hydroxybenzoate-grown *P. putida* 90.

detectable activities of either muconate lactonizing enzyme or muconolactone isomerase, and likewise do not cross-react with either antiserum (Fig. 7).

The failure of the extract from *p*-hydroxybenzoate-grown cells to cross-react is of considerable interest, in view of earlier enzymological findings (Ornston, 1966a, b). In the protocatechuate branch of the pathway, carboxymuconate lactonizing enzyme and carboxymuconolactone decarboxylase catalyse reactions that are chemically analogous to those catalysed in the catechol branch by muconate lactonizing enzyme and

muconolactone isomerase, respectively. Ornston found that carboxymuconate lactonizing enzyme and muconate lactonizing enzyme, as well as carboxymuconolactone decarboxylase and muconolactone isomerase, constitute chemically similar pairs of proteins in terms of gross chemical criteria: molecular weight, thermal stability, pH optimum. Coupled with their catalytic resemblances, this led him to suggest that two pairs might be related in primary structure, having arisen from common protein precursors in the evolution of the species. However, the immunological findings provide no support for this hypothesis.

These two antisera were tested qualitatively for their ability to cross-react with extracts from appropriately induced cells of other strains and species. On each gel diffusion plate, two antigen wells were furnished with extracts of the homologous strain, prepared from cells grown on benzoate and succinate, respectively. Adjacent antigen wells were furnished with extracts of the strain to be tested, grown on the same two substrates. In all cases, antigen wells furnished with extracts from benzoate-grown cells contained equal unitages of enzyme, and antigen wells furnished with extracts from succinate-grown cells contained equivalent amounts of total protein. The extracts of succinate-grown cells never yielded precipitin lines with either antiserum. They accordingly provided an extremely useful control, since if cross-reactions did occur with extracts of benzoate-grown cells, it was immediately evident that the cross-reacting materials had been specifically induced as a result of growth on benzoate.

Before presenting the results of these experiments, I must digress briefly to describe presumed taxonomic relationships within the genus *Pseudomonas*, recently studied in great detail by Stanier *et al.* (1966), Redfearn *et al.* (1966) and Mandel (1966). *Pseudomonas putida* is a member of a subgeneric cluster known as the fluorescent group, distinguishable from other *Pseudomonas* spp. by many common phenotypic properties. Within this cluster, we have recognized three species. *Pseudomonas aeruginosa* is internally very uniform, and easily distinguishable from the other two species by many phenotypic traits, as well as by DNA base composition. *Pseudomonas putida* and *Pseudomonas fluorescens* are far less uniform internally, and the phenotypic distinctions between them are relatively few in number. The internal heterogeneities led us to recognize two biotypes of *P. putida* (A and B) and no less than seven biotypes of *P. fluorescens* (A to G). The biotypes differ from one another in minor phenotypic respects, and also in some cases by small differences in DNA base composition. Some of the characters that distinguish species and biotypes in the fluorescent group are listed in Table 3. Outside the fluorescent group, the β-ketoadipate pathway occurs in two phenotypically isolated *Pseudomonas* species,

TABLE 3. Some differences between species and biotypes of fluorescent pseudomonads

	P. aeruginosa	*P. putida*, biotype		*P. fluorescens*, biotype				
		A	B	A	B	C	D	E
Moles % G + C in DNA	67·2 ± 1·1	62·5 ± 0·9	60·7 ± 1·1	60·5 ± 1·1	61·3 ± 1·1	60·6 ± 1·1	63·5 ± 0·9	63·6 ± 0·8
No. of flagella	1	> 1	> 1	> 1	> 1	> 1	> 1	> 1
Gelatinase production	+	−	−	+	+	+	+	+
Egg-yolk reaction	−	−	−	+	+ or −	+	+	+
Phenazine pigments	+P[+]	−	−	−	−	−	+C[+]	+pc[+]
Denitrification	+	−	−	−	+	+	+	−
Growth at 4°C	−	+ or −	+	+	+	+	+	+
41°C	+	−	−	−	−	−	−	−
Use as sole source of carbon and energy:								
L-arabinose	−	+ or −	+	+	+	−	−	+
trehalose	−	−	−	+	+	+	+	+
sucrose	−	−	−	+L[a]	+L[a]	−	+L[a]	+L[a]
m-inositol	−	−	−	+	+	+	+	+
geraniol	+	−	−	−	−	−	−	−
acetamide	+	+ or −	+ or −	−	−	−	−	−
propylene glycol	+	+ or −	−	−	+	+	−	−
L-tryptophan	+ or −	−	+	+	+ or −	+	+	+
glycine	+	+ or −	+	−	−	−	−	−
benzylamine	−	+	+	−	−	−	−	−

[a] L denotes levan formation. +P = pyocyanin; C = chlororaphin; pc = phenazine-α-carboxylic acid.

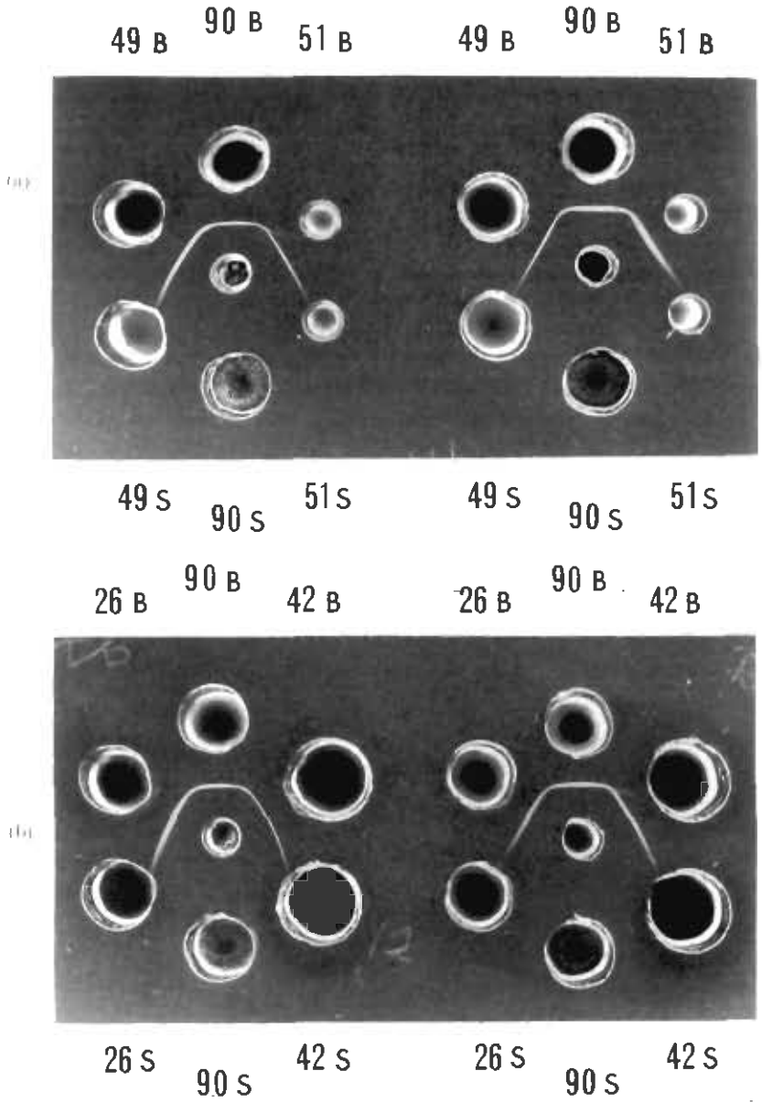


FIG. 8. Precipitin reactions of several strains of *P. putida* biotype A with anti-isomerase (left centre well) and anti-lactonizing enzyme (right centre well). Other wells contained crude extracts of cells grown either with benzoate (suffix B) or succinate (suffix S); the numbers designate strains; 90 is the strain from which the immunizing antigens were prepared. Note that two additional strains show complete antigenic homology with 90, one shows a faint spur against anti-isomerase, and one a faint spur against anti-lactonizing enzyme.

P. multivorans and *P. stutzeri*, as well as in the pseudomallei species group (*P. pseudomallei* and *P. mallei*).

The immunological point of reference in our work was a strain of *P. putida* biotype A. Extracts from benzoate-grown cells of other strains belonging to *P. putida* biotype A cross-reacted with both antisera in a manner which indicated almost complete homology. Certain of these extracts showed faint spurring, sometimes with only one of the antisera, sometimes with both (Fig. 8). These observations nicely confirm the internal uniformity of *P. putida* biotype A. Extracts of *P. aeruginosa* cross-reacted strongly with both antisera; but heavy spurs were always

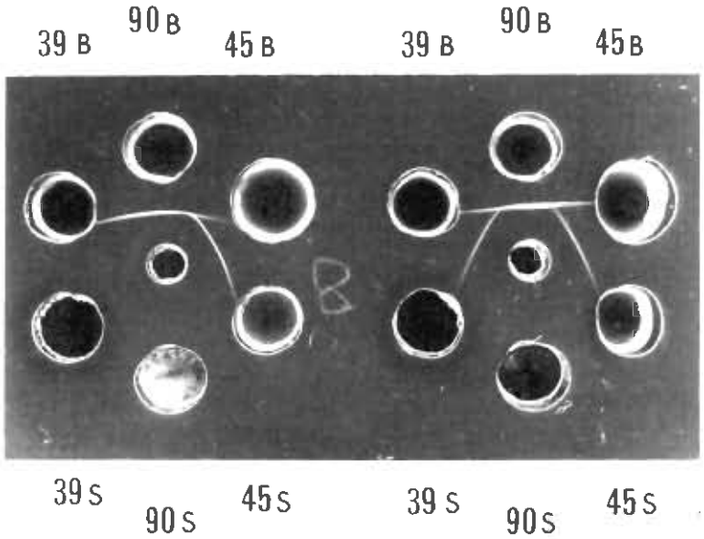


FIG. 9. Precipitin reactions of a strain of *P. fluorescens* biotype E and a strain of *P. aeruginosa* with anti-isomerase (left centre well) and anti-lactonizing enzyme (right centre well). All other wells contained crude extracts of cells grown with benzoate (B) or succinate (S). 90: *P. putida* 90 (homologous strain), 39: *P. fluorescens* biotype E; 45: *P. aeruginosa*.

evident, indicative of only partial antigenic homology (Fig. 9). Extracts of strains belonging to every biotype of *P. fluorescens* gave cross-reactions with heavy spurring against anti-lactonizing enzyme, and exceedingly weak and diffuse cross-reactions against anti-isomerase. The heterologous reactions on gel diffusion plates are so characteristic that they could easily serve for specific identification, as shown in Figs. 9 and 10. Immunological criteria have long been used in bacterial

classification. However, the traditional serotaxonomy of bacteria has rested almost exclusively on the characterization of the surface antigens of the cell, i.e. those located in the wall, the flagella and fimbriae (pili), and the capsules. As a general rule, such analyses have proved of limited value for the recognition of species; they cut too fine, and are mostly of use to discriminate between strains belonging to a single species. This obviously suggests that the surface antigens of the bacterial cell are subject to very rapid evolutionary change. Our data imply that the enzymes of the bacterial cell may be far more stable,

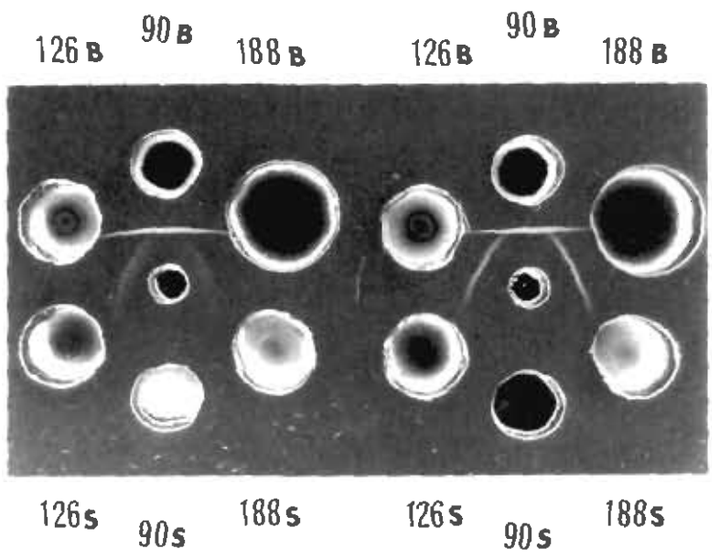


FIG. 10. Precipitin reactions of two strains of *P. fluorescens* biotype A with anti-isomerase (*left centre well*) and anti-lactonizing enzyme (*right centre well*). All other wells contained crude extracts of cells grown with benzoate (suffix B) or succinate (suffix S). 90: *P. putida* 90 (homologous strain); 126 and 188: two strains of *P. fluorescens* biotype A. Note the great difference in the strength of heterologous reactions with the two anti-enzymes.

evolutionarily speaking, and therefore provide more suitable antigenic markers for specific differentiation.

The results obtained with extracts from strains of *P. putida* biotype B revealed that we had erred in associating these strains taxonomically with *P. putida* biotype A. The gel diffusion patterns closely resembled those obtained with the various biotypes of *P. fluorescens*, except that

cross-reactions with the anti-isomerase were even weaker, in a few cases undetectable. A different taxonomic allocation for these strains accordingly seems called for.

Cross-reactivity with the two antisera extends beyond the fluorescent group to include *P. stutzeri*, extracts of which react weakly with anti-isomerase, and more strongly with anti-lactonizing enzyme (Fig. 11). *However, we have been unable to detect cross-reactions* with either antiserum by extracts of *P. multiroraus*. The members of the pseudomallei group have not yet been tested. Extracts of *Hydrogenomonas eutropha* and *Moraxella calcoacetica* do not cross-react. The combined

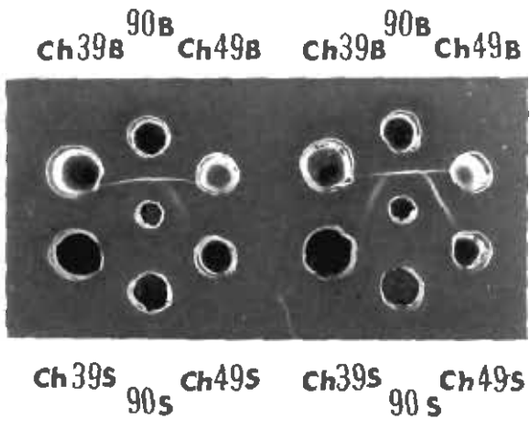


FIG. 11. Precipitin reactions of two strains of *P. stutzeri* with anti-isomerase (left centre well) and anti-lactonizing enzyme (right centre well). All other wells contained crude extracts of cells grown with benzoate (suffix B) or succinate (suffix S). 90: *P. putida* 90 (homologous strains); CH 39 and CH 49: two strains of *P. stutzeri*.

biochemical and immunological findings suggest the hypothesis concerning the evolution of the β-ketoadipate pathway presented in Fig. 12.

Using the technique of microcomplement fixation, we have estimated quantitatively the antigenic relationship for a few strains belonging to each of the cross-reacting species and biotypes. The results are summarized in Table 4, antigenic relatedness being numerically expressed in terms of an "index of dissimilarity".

The index of dissimilarity expresses the factor by which the concentration of antiserum must be raised in order that a particular cell-free extract containing enzyme gives a peak height equal to that

TABLE 4. *Quantitative immunological relationships of two enzymes of the β-ketoadipate pathway among Pseudomonas spp., determined with antisera directed against pure enzymes isolated from a strain of P. putida biotype A*

Species and biotype	Number of strains examined	Index of dissimilarity[a]	
		Muconate lactonizing enzyme	Muconolactone isomerase
P. putida biotype A	7	0·95–2·25	1·0–1·6
P. putida biotype B	6	13–17	> 120
P. fluorescens biotype A	3	12–20	> 120
P. fluorescens biotype B	3	21·5–22	> 120
P. fluorescens biotype C	3	34–60	> 120
P. fluorescens biotype D	3	10·5–20·5	> 120
P. fluorescens biotype E	3	10·5–16	> 120
P. aeruginosa	3	11–12	2·8–3·0
P. stutzeri	5	10–21	100–> 120

[a] A value of 1 indicates antigenic identity; values greater than 1 indicate increasing degrees of dissimilarity.

yielded by a cell-free extract containing the enzyme used as immunizing antigen. A value of 1·0 indicates complete antigenic identity; values greater than 1·0 indicate increasing degrees of dissimilarity.

The quantitative data reveal many interesting facts not evident from the qualitative experiments. In the first place, different strains of each species or biotype examined show reasonable constancy with respect to the antigenic properties of the two enzymes examined. Secondly, although muconate lactonizing enzyme and muconolactone isomerase are closely linked in terms of both synthesis and function, it is evident that they have evolved at markedly different rates in the various species of the *Pseudomonas* group. As a rule, antigenic divergence from the reference strain has been more pronounced for the isomerase than for the lactonizing enzyme, although this rule does not hold for *P. aeruginosa*. Thirdly, it can be seen that a substantial divergence between two species with respect to the base content of the DNA does not necessarily result in modifications of protein structure that abolish antigenic homology. *P. putida* biotype A and *P. aeruginosa* differ by nearly 5 moles % in the guanine + cytosine contents of their DNA; but both enzymes retain relatively high antigenic homology between the two species.

When compared with similar immunological data for enzymes of vertebrates, these data begin to show us just how different in evolutionary terms the "species" of the bacteriologist and the zoologist

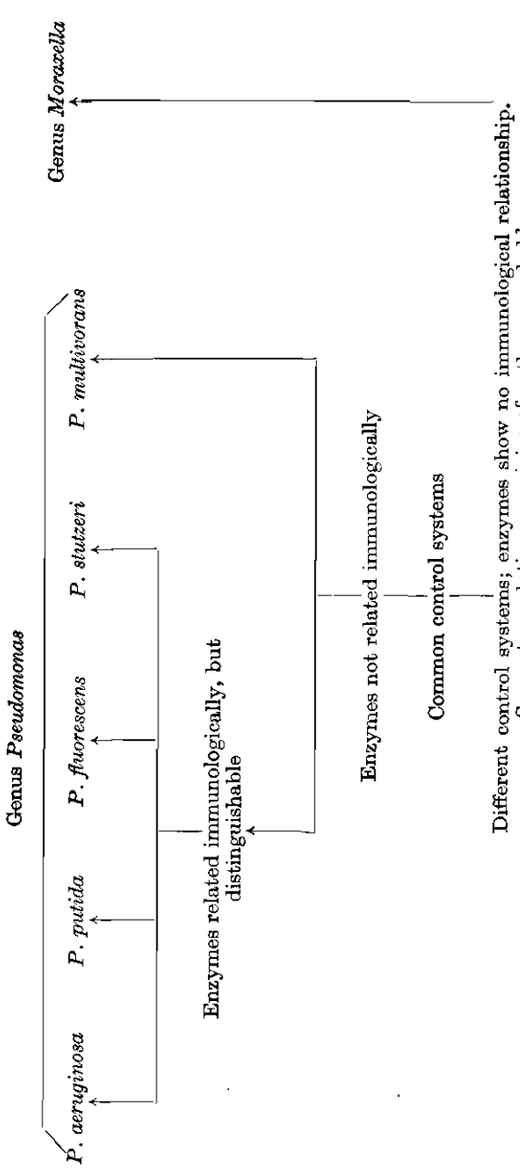


FIG. 12. An interpretation of the evolution of the β-ketoadipate pathway in two bacterial groups.

TABLE 5. Quantitative immunological relationships of several enzymes among vertebrates, determined with *antisera* to pure enzymes of the chicken. Data of Wilson *et al.* (1964)

Class	Species	Index of dissimilarity			
		Heart lactic dehydrogenase	Muscle lactic dehydrogenase	Triose-P dehydrogenase	Aldolase
Aves	Chicken	1·0	1·0	1·0	1·0
	Turkey	1·4	1·2	1·0	1·0
	Duck	1·5	4·3	1·2	
	Ostrich	1·9	3·1	1·3	5·0
Reptilia	Caiman	3·3	4·2	3·8	6·5
	Painted turtle	4·0	5·2	4·2	
Amphibia	Bullfrog	14	40	30	18
Pisces	Sturgeon	80	20	12	25
	Halibut		> 200	> 50	> 100
	Dog fish	> 100	> 200	> 50	
	Lamprey		30	> 50	

really are. In Table 5, I have summarized some of the data of Wilson *et al.* (1964), also obtained by the technique of micro-complement fixation, on enzymes from vertebrates. In these experiments, the chicken served as the immunological reference point. It is necessary to look beyond the classes Aves and Reptilia, to the Amphibia, in order to find indices of dissimilarity as great as those that separate *Pseudomonas putida* from *Pseudomonas fluorescens*. Yet as I have pointed out previously, these two bacterial species are very similar in gross phenotype, and their taxonomic differentiation is not easy. This comparison between vertebrates and bacteria must not, of course, be taken too literally. The data do not necessarily imply that *P. putida* and *P. fluorescens* are separated by the same *absolute time span* as that which separates the bullfrog from the chicken, since the tempo of protein evolution is probably much more rapid in bacteria than in vertebrates.

It can be concluded that despite the inherently difficult biological material with which he works, the bacteriologist does have at his disposal biochemical and immunological methods that can be used to explore effectively problems of evolutionary affinity at the lower end of the contemporary biological scale.

ACKNOWLEDGMENT

I am indebted to Dr. Allan C. Wilson for his collaboration in the immunological studies, and to Mrs. Charlotte Gasser and Mr. Douglas Wachter for technical assistance in carrying out the experiments.

This work was supported by a U.S.H.P.S. grant from the National Institutes of Health (AI-1808) and by a National Science Foundation grant (GB-4112).

REFERENCES

Cánovas, J. L. and Stanier, R. Y. (1967). Regulation of the enzymes of the β-ketoadipate pathway in *Moraxella calcoacetica*. I. General aspects. *Eur. J. Biochem.* **1**, 289–300.

Cohn, F. (1872). Untersuchungen uber Bacterien. *Beitr. Biol. Pfl.* **2**, 127–224.

Hill, L. R. (1966). An index to deoxyribonucleic acid base composition of bacterial species. *J. gen. Microbiol.* **44**, 419–37.

Mandel, M. (1966). Deoxyribonucleic acid base composition in the genus *Pseudomonas*. *J. gen. Microbiol.* **43**, 273–92.

Ornston, L. N. (1966a). The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida*. III. Enzymes of the catechol pathway. *J. biol. Chem.* **241**, 3795–99.

Ornston, L. N. (1966b). The conversion of catechol and protocatechuate to β-ketoadipate by *Pseudomonas putida*. IV. Regulation. *J. biol. Chem.* **241**, 3800–10.

Postgate, J. R. (1965). Recent advances in the study of the sulfate reducing bacteria. *Bact. Rev.* **28**, 425–41.

Redfearn, M. S., Palleroni, N. J. and Stanier, R. Y. (1966). A comparative study of *Pseudomonas pseudomallei* and *Bacillus mallei*. *J. gen. Microbiol.* **43**, 293–313.

Stanier, R. Y. (1964). Towards a definition of the bacteria. *In* Gunsalus, I. C. and Stanier, R. Y. (*eds.*) "The Bacteria" Vol. 5, **445–64**. Academic Press, New York.

Stanier, R. Y., Palleroni, N. J. and Doudoroff, M. (1966). The aerobic pseudomonads: a taxonomic study. *J. gen. Microbiol.* **43**, 159–272.

Stanier, R. Y. and van Niel, C. B. (1962). The concept of a bacterium. *Archs Mikrobiol.* **42**, 17–35.

Wilson, A. C., Kaplan, N. O., Lurine, L., Pesee, A., Reichlin, M. and Allison, W. S. (1964). Evolution of lactic dehydrogenases. *Fed. Proc. Fedn Am. Socs exp. Biol.* **23**, 1258–66.

Note added in proof: The generic name *Moraxella* has been applied in this paper to bacteria related to the species first described by Beijerinck as *Micrococcus calco-aceticus* and to organisms known as *Bacterium anitratum*, *Moraxella lwoffi*, etc. We have recently proposed the classification of this group in the genus *Acinetobacter* (P. Baumann, M. Doudoroff, and R. Y. Stanier, *Journal of Bacteriology*, **95**, 1520–41 (1968)).

SECTION V

THE ASSESSMENT OF BIOCHEMICAL TECHNIQUES IN TAXONOMY

19. The Assessment of New Types of Character in Taxonomy

A. J. Cain

Department of Zoology, University of Manchester, England

Before one can talk about the assessment of the taxonomic value of new chemical, biochemical, serological or in fact any techniques, one must be fairly clear about what is meant by taxonomic value, and what is the value of the so-called old or classical taxonomy. Some biochemists perhaps treat classical taxonomy with rather more than necessary deference; others seem convinced that whatever their findings are, and on however narrow a basis (presence or absence of a single chemical compound, single specimens of single species taken as adequate samples of a whole order, class or phylum) the old taxonomy is superseded since their epoch-making labours. Such attitudes are less common than they were twenty years ago (they are in fact very old attitudes) and most people probably take a stand somewhere between these two extremes. Taxonomists are largely to blame for these attitudes. Even now there are a number that work on in the old traditional way with little comprehension of the reasons for it, and though there are several very useful books (for example Blackwelder, 1967; Mayr *et al.*, 1953; Sokal and Sneath, 1963; Simpson, 1961; Davis and Heywood, 1963) there is still no satisfactory account of taxonomic principles. I am not going to try to produce one in the space of this short paper, only to indicate some of the most important features.

First of all the point should be made that although the "old" taxonomy is traditionally based on the examination and dissection of preserved material, this is in fact a very wide basis. If there were in each organism one pair of chromosomes carrying all the genes for the anatomy, another for those involved in digestive processes, yet another for those controlling reproduction, and so on, it might well be that to look only at anatomy might give a very insufficient sampling of the

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 229–234.

genotype; as things are, there is no trace of such a limitation and it is likely that a classification based on a really careful study of anatomy is a better one than any on, say, proportions of two different lipids in the central nervous system. Such classifications on very restricted bases have been known for a long time. Darwin (1859), whose treatment of the whole subject in his Chapter 13 is masterly, puts it very clearly: "The importance, for classification, of trifling characters, mainly depends on their being correlated with several other characters of more or less importance. The value indeed of an aggregate of characters is very evident in natural history. Hence, as has often been remarked, a species may depart from its allies in several characters, both of high physiological importance and of almost universal prevalence, and yet leave us in no doubt where it should be ranked. Hence, also, it has been found that a classification founded on any single character, however important that may be, has always failed; for no part of the organization is invariably constant." This is perfectly true. Except in devising classifications for the purpose of identification, i.e. keys, all the characters that can be examined must be taken into account. Unfortunately, since the work of museum taxonomists contains a great deal of identification, and this is greatly facilitated when good short snappy, easily used, definitions can be given of the groups in question, there is a strong tendency to erect "natural" groups which are at least in part really key groups, and a truly natural classification is not then produced.

The word "natural" in relation to classification has been used in many different ways (Cain, 1954, 1959c, d, 1962a, b, 1963; Cain and Harrison, 1958, 1960). It means here a classification based on many characters as against one on few, the latter being an artificial or a key classification. In the last century, however, it has usually been used to mean a classification in accordance with the actual course of evolution—a phylogenetic or *phyletic* classification. Such has been the prestige of the theory of evolution that many have felt compelled to erect the most remarkable family trees on the least possible evidence—often selecting variation in only one particular sort of characters as the basis because, on very dubious principles, it was thought to be of greater evolutionary importance. It cannot be too strongly emphasized that the *only* secure basis for a phyletic classification is a good fossil record; moreover, the successive stages in a given phyletic line are recognized by putting together those specimens found which are in all respects most like each other, in other words by a natural classification in the sense given above. To contrast this natural classification with phyletic classifications made on some special basis, Cain and Harrison (1960) called it *phenetic* classification or arrangement, because it uses without weighting any

sort of available character that appears in the material, including, of course, genetic characters. Incidentally, it is a striking example of the prostitution of terms that already phenetic is being used in taxonomic contexts as a mere abbreviation of phenotypic. Where only a poor fossil record is available, as in most groups of the animal kingdom, a phenetic classification is the only possible "natural" one. Often the groups found can be arranged in a sort of dendrogram, but it must be pointed out that a dendrogram, especially one based only on modern forms, is not a family tree and cannot be, except by the most unlikely coincidence.

In making comparisons between different forms (or specimens) it is essential to compare only homologous characters, as Woodger has pointed out (see Cain and Harrison, *1958 for discussion*) and the material is arranged so that there is the maximum number of homologous correspondences between the specimens in hand. In taxonomy we do not start comparing one lizard from the snout backwards with another examined from the tip of the tail forwards (though this might be useful for some special purpose). Erdtman (this volume) has raised a most important point here. He shows that some very complex compounds may be synthesized by different pathways and should then not be regarded as homologous in different organisms unless the pathways are shown to be the same. But in fact even this is not sufficient. Haemoglobins occur sporadically in the most diverse groups of the animal kingdom; cytochromes are an exceedingly widespread if not universal constituent of living matter and the derivation of haemoglobins and similar compounds from them by the *same* enzymic pathways may therefore have occurred independently many times. It is impossible that a few freshwater snails, some *Chironomus* gnat larvae (but not all), one or two genera of marine annelids, and the vertebrates are a natural group because they all contain haemoglobin, and still less could we include with them root nodules of some legumes which also (as I learnt from our Chairman) contain haemoglobin. Here is a fine example of the absolute necessity of checking against the "old" taxonomy before coming to conclusions. The same could be said for as different a character as chromosome number (Cain, 1958a).

What does actually govern the distribution of a particular character in living things? I have given reasons elsewhere (Cain, 1959d, 1964) for believing that the power of natural selection is far greater than has been thought before; that, with a lag which is appreciable but trivial on the scale of geological time, characters are evolved as they are wanted; that the very general explaining away of characters as due to their occurrence in remote ancestors is simply based on the true fact that the author indulging in it does not know anything about the selective forces bearing on the characters concerned; and that ignorance is a poor foundation for

a very definite conclusion. On the positive side it is certainly true that whenever the most apparently trivial or neutral or non-functional or ancestral characters have in fact been investigated, it has been found either that they have a very definite function or at least that selection coefficients greater than what is necessary to dominate their distribution are in fact acting on them in the wild. The papers in the symposium on Function and Taxonomic Importance published by the Systematics Association, especially that by Dr. S. Manton, make this point, as do many others referred to by Cain (1964). But the moment we accept this point we are faced with considerable possibilities of convergence (Cain 1959a, b, d). If a particular structure or compound is needed, it will be evolved as in the case of the haemoglobins mentioned above—and Huntsman (this volume) shows beautifully how even the fine details of the haemoglobin molecule are highly functional and therefore subject to intense selection (even more in the wild than in the laboratory). If a structure or compound is highly functional, then it is likely to be as we find it because it is needed, not at all merely because an ancestor had it; its present distribution, therefore, need give us no clue at all to the course of evolution. When far-reaching changes in the accepted view of the relationships of major groups of vertebrates seem to be implied by biochemical or serological studies (e.g. Blombäck and Blombäck, Gell, this volume), especially when they seem to disagree with a fossil record of some importance, it is worth while stopping and asking whether similarities in large molecules have with absolute certainty no functional explanation at all but are pure potted history. Equally, when Haslewood (this volume) in the course of a fascinating review of bile salts in vertebrates suggests that there is some connection between certain salts and a vegetarian diet, we immediately wonder what other associations between a particular diet (and possibly the medium in which it is engulfed?) and a given chemical structure may exist, or how far a particular bile salt must be fitted in with other bile constituents to be effective. All demonstrations of possible function mean that we can no longer simply interpret resemblances as historical without giving independent evidence that this is indeed the case. A great deal of the development of taxonomy (Cain 1958b, 1959a, b, 1962a, b) has been through the eliminating of principles (logical, physiological, and now evolutionary) that have been used to produce unwarranted conclusions; and the principle that any character must show in its variation the course of history if on a casual inspection no function for it can be seen, is at least as invalid as any of its predecessors.

No doubt if we could get at the actual arrangement of the genetic material of each organism, we might in some sense be able to make a more fundamental classification. But even here, I wonder if the genetic

arrangements have not themselves come under selection. Suppose that we knew for any form the whole genetic coding and could decipher it. It might be the best shorthand description of the whole phenotype. Work with this in mind is intensely exciting. It should be pushed forward as vigorously as possible. But there is some danger of expecting too much from it for taxonomy. In the first place the extinct animals seem beyond its reach, yet we cannot omit them from our classification. Secondly, we may find convergence there as much as elsewhere. At the present day the remarkable methods described by Kohne (this volume) do seem to allow us to get some value of overall resemblance between two forms; and anything that will do this will be welcome to the taxonomist who does not relish the prospect, even with computer aid, of having to examine many hundreds or thousands of characters in every specimen in order to get a good phenetic arrangement. But so far we have been thinking of positive comparison. Kohne's methods do indeed give us some indication of at how many points along two strands identical bases occur—that is, there is some measure of amount of identity. But where there is no pairing, all we know is that the bases are different, and as there are more than two alternatives at each site, we may be in some difficulty here. These methods tell us the amount of identity (disregarding convergence) but not the amount of difference. To caricature this situation in order to make it clearer, one might be told that two animals were identical (at the accuracy of investigation) in one character, but in all the rest all one could say was that one was not the other. It is a situation not unknown in orthodox taxonomy, in which I know one or two groups whose only real definition is that they are not members of some other group.

All that the really competent taxonomist needs is to know everything about his subject and about the groups he works on. Since this is hardly possible he must go to others for help. The new taxonomy is peculiarly a field in which team work is necessary, and as this conference has so well shown, team work by taxonomists, chemists, biochemists and anyone else with special knowledge can be stimulating, profitable and delightful.

REFERENCES

Blackwelder, R. E. (1967). "Taxonomy." Wiley, New York.

Cain, A. J. (1954). "Animal Species and their Evolution." Hutchinson, London.

Cain, A. J. (1958a). Chromosomes and taxonomic importances. *Proc. Linn. Soc. Lond.* **169**, 125–8.

Cain, A. J. (1958b). Logic and memory in Linnaeus's system of taxonomy. *Proc. Linn. Soc. Lond.* **169**, 144–63.

Cain, A. J. (1959a). Deductive and inductive methods in post-Linnaean taxonomy. *Proc. Linn. Soc. Lond.* **170**, 185–217.

Cain, A. J. (1959b). The post-Linnaean development of taxonomy. *Proc. Linn. Soc. Lond.* **170**, 234–44.

Cain, A. J. (1959c). Taxonomic concepts. *Ibis* **101**, 302–18.

Cain, A. J. (1959d). Function and taxonomic importance. *In* Cain, A. J. (ed.) "Function and Taxonomic Importance". *Systematics Association*, London.

Cain, A. J. (1962a). The evolution of taxonomic principles. *Symp. Soc. gen. Microbiol.* **12**, 1–13.

Cain, A. J. (1962b). Zoological classification. *Aslib Proc.* **14**, 226–30.

Cain, A. J. (1963). The natural classification. *Proc. Linn. Soc. Lond.* **174**, 115–21.

Cain, A. J. (1964). The perfection of animals. "Viewpoints in Biology" (3), 36–63.

Cain, A. J. and Harrison, G. A. (1958). An analysis of the taxonomist's judgement of affinity. *Proc. Zool. Soc. Lond.* **131**, 85–98.

Cain, A. J. and Harrison, G. A. (1960). Phyletic weighting. *Proc. Zool. Soc. Lond.* **135**, 1–31.

Darwin, C. (1859). "On the Origin of Species." Murray, London.

Davis, P. H. and Heywood, V. H. (1963). "Principles of Angiosperm Taxonomy." Oliver and Boyd, Edinburgh and London.

Mayr, E., Linsley, E. G. and Usinger, R. L. (1953). "Methods and Principles of Systematic Zoology." McGraw-Hill, New York.

Simpson, G. G. (1961). "Principles of Animal Taxonomy." Columbia University Press, New York.

Sokal, R. R. and Sneath, P. H. A. (1963). "Principles of Numerical Taxonomy." Freeman, San Francisco.

20. The Assessment of Biochemical Techniques in Plant Taxonomy

HOLGER ERDTMAN

Department of Organic Chemistry, Royal Institute of Technology, Stockholm, Sweden

ABSTRACT

Chemical constituents of organisms are taxonomically valuable and have the advantage over classical characters that they can be exactly described. At the present time they are not as useful in some taxa as in others but most of the chemical screening of taxa still remains to be done. Their advantage is somewhat reduced by the fact that chemically identical compounds may be biosynthetically different. Moreover, the use of chemical characteristics is essentially limited to extant species.

Whenever possible, constituents occurring in homologous organs in different species should be compared. In molecular taxonomy, patterns of compounds, preferably covering a range of biosynthetic origins, are much more useful than solitary compounds. The elucidation of the biosynthesis of compounds used for taxonomic purposes is vitally important for the further development of molecular taxonomy.

INTRODUCTION

In the areas where sciences overlap there can be considerable difficulties in communication between different specialists. The areas where biological systematics and organic chemistry merge are no exception. However, it is my experience that biologists find it easier to understand organic chemists than *vice versa*. The results of structural organic chemistry can be presented in simple constitutional symbols and the relationships between these are intelligible to most biologists, probably because of their training in comparative anatomy and morphology.

To chemists, on the other hand, systematic biology mostly appears to be an enormous collection of more or less unrelated facts. Their understanding of the evolutionary changes which life has undergone is

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 235–268.

hampered by their lack of appreciation of the length of time required for the earliest forms of living organisms to give rise to the present flora and fauna.

TAXONOMY AND CHEMISTRY

General remarks

Our theories and ideas about evolution during the "biological" era, the last 500 million years, are the result of the labour and the constructive ideas of a very large number of different scientists, particularly botanists, zoologists, geneticists, geologists and palaeogeographers. However, as is reflected by the different classifications that are proposed from time to time there is still no definitive understanding of phylogenetic relations.

The unravelling of the evolutionary processes is one of the most difficult problems that natural science has faced. Present views, particularly with regard to plants, have been admirably summarized by Stebbins (1951, p. 513). He writes: "Mutation, recombination, and selection, as well as isolation and extinction, are responsible for groups and categories at all levels. As we go higher up the scale, reproductive isolation becomes less important as the basis for the limitation of categories, and extinction more so, but this is a quantitative rather than a qualitative change. Once species have appeared, the origin of genera and families, as well as the orders, classes and phyla, is largely a matter of time and further genetic plus environmental change."

However, the genetic tree, so popular among non-biologists, and for that matter also with some biologists, is a model of doubtful value. A more realistic picture is probably that of a group of bushes with spreading branches. The branches themselves might be compared with bundles of interlacing fibres. The composition of the present flora and fauna would be represented by the growing tips of the branches.

At a generic and frequently also at a family level we may be right in speaking about monophyletic relations, but on the level of orders, classes etc. phylogenetic relations are unknown or at least very speculative.

It may appear surprising that chemistry has played only a very minor role in taxonomy and phylogeny, particularly since it is obvious that, for example, morphogenesis has chemical causes and that, ultimately, it should be possible to describe the ontogenesis of an organism in chemical terms. Moreover, chemical characters have the great advantage that they can be described exactly by molecular structures and configurations. However, unfortunately, chemical constituents cannot be used for the characterization of fossils except in very rare cases.

The lack of contributions from chemistry is, of course, conditioned in

part by the later development of organic chemical techniques. Until recently the structural elucidation of even fairly simple organic molecules was a long, laborious and costly affair. Many groups of compounds of great general interest have long remained unknown simply because the substances did not exhibit any conspicuous physical or physiological properties or because they did not readily crystallize from crude extracts.

The recent introduction of new and powerful chemical, physical and analytical methods has completely changed the situation and has made the isolation and structural elucidation of natural products very much easier. The number of substances of known structure is therefore at present increasing at a fantastic rate.

The spectacular advances in biochemistry and molecular biology and the increasingly clear picture of the structural and configurational relations between different groups or "families" of natural products are giving rise to an interest in biological problems among a growing number of organic chemists. There are two main fields of current interest in biological organic chemistry, the biosynthesis of natural products and the investigation of the use of chemical characteristics in taxonomy.

Chemotaxonomy? Biochemical systematics? Molecular taxonomy?

The hybrid science "chemotaxonomy" has passed through a long induction period, but the "sterility barriers" caused by the divergent development of biology and chemistry are now steadily being broken down. The term chemotaxonomy is old and like many similar terms rather loose. Some authors prefer "biochemical systematics" or even "chemical systematics" though the latter term is hardly acceptable to organic chemists who have their own systematic problems. Other authors avoid the words systematics and taxonomy altogether using instead comparative phyto- or zoo-chemistry but this is to put too much emphasis on the chemical side of the problem. The discussion is obviously due to the fact that taxonomy and systematics are used as synonyms in some circles whereas in others they are differently defined. Personally I prefer the term molecular taxonomy because what we are dealing with is very definitely molecules in relation to taxonomy (or systematics). However, it matters little which term is used as long as there is agreement on what it means.

CHEMICAL CHARACTERISTICS

Primary and secondary products

Some of the molecules of which a plant is composed are high-molecular weight compounds serving structural purposes, constituting nutritional reserves or being components of the enzymatic machinery

of the cells. The low-molecular weight substances are frequently divided into primary and secondary products but it is impossible to give an exact definition of these terms. The primary products are those involved in the fundamental metabolic processes and they are normally not accumulated. The secondary products however are often stored, sometimes in large amounts, within the living cells or may be deposited in glands or resin ducts or in tissues such as the bark or heartwood. They are often "metabolic by-products" and are particularly common in plants, which do not have the same possibilities as animals for getting rid of waste products. Some of these products are not definitively expelled from the metabolism and can undergo further changes. This sometimes gives rise to diurnal or seasonal variations which occasionally have led chemists to draw false conclusions from analytical investigations.

The question of whether the secondary products have, or have had, any function has been much discussed but clearly many of them are of considerable value to the organisms producing them. Among these are the colouring matters of flowers, insect repellent or attracting compounds, pheromones, the waxy exudates from the preen glands of certain birds and the cholic acids of vertebrates. Poisonous or bitter plants are mostly avoided by animals but usually such compounds are not essential for the life of the plant and cultivated races of such species sometimes do not contain any of these obnoxious principles. However, whether all these substances can be properly included among the so-called secondary products is debatable.

Because of their availability and economic value, many of the secondary products of metabolism have long aroused the interest of the organic chemist. Indeed the structural problems they present and the methods required to solve them have been a basic theme in the development of a large part of organic chemistry.

There is a natural tendency for the organic chemist to regard the secondary products as particularly important from a taxonomic point of view but this is certainly wrong. The primary processes and the enzymatic and other factors governing them will probably prove to be of more fundamental value in taxonomy. However, we still know too little about them to be able to include them in molecular taxonomic discussion. For the time being, therefore, molecular taxonomy largely rests on the biochemically less interesting secondary products.

Families of chemical constituents

The observation that taxonomically related plants often contain identical or at least closely related compounds is the very foundation of molecular taxonomy. Recognition of this has practical consequences, and for instance the purposeful search for certain useful compounds in

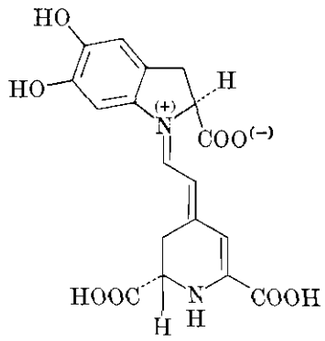
related taxa has been of great interest to the pharmaceutical and food industries.

Chemists, particularly in the alkaloid field, have recognized a number of groups of compounds more or less characteristic of particular families and talk about "Senecio alkaloids", "Lupin alkaloids", "Amaryllidaceae alkaloids", "Lycopodium alkaloids" and so on. However, there are very many taxa lacking such chemical characterization, at least at present. Sometimes identical or chemically closely related secondary products occur in hiologically quite unrelated organisms. Examples are nicotine in *Nicotiana* and *Equisetum*, anabasine in *Nicotiana* and *Anabasis*, bufotenine in the toad, *Bufo*, and in *Piptadenia* (Leguminosae). The sesquiterpene dendrolasin produced by *Lasius*, an ant, for defence purposes also occurs in the wood of *Torreya nucifera*. Even such a common terpene as α-pinene is used by arthropods (*Nasutitermes*) as a repellent. Facts of this kind have been used as arguments against molecular taxonomy. Parallel development or convergence are accepted phenomena in phylogeny. Why should not similar phenomena be acceptable in molecular taxonomy?

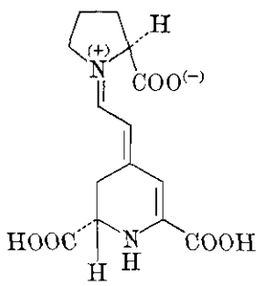
Distribution of chemical characteristics and their taxonomic value

Compounds which are ubiquitous or nearly so are obviously of little taxonomic value and thus many biochemically important substances are taxonomically useless. Nobody would deny that compounds such as cellulose, lignins, chlorophyll etc. have taxonomic interest but their importance is limited to the delineation of higher categories.

With the exception of the suborder Caryophyllineae all plants belonging to the order Centrospermae (including Cactales) appear to produce nitrogenous colouring matters instead of anthocyanins. Examples are betanidin (1) from red beet and indicaxanthin (2) from *Opuntia ficus*



(1) Betanidin

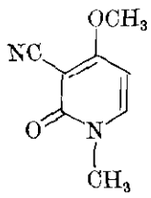


(2) Indicaxanthin

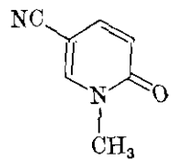
indica. These "betacyanins" contain a characteristic pyridine moiety combined with amino acids. The anthocyanins belong to the class of flavonoids and are common colouring matters of the flowers of most angiosperms including Caryophyllineae. They are also found in gymnosperms, ferns and some mosses. The lack of anthocyanins in the betacyanin-producing plants is interesting particularly because they are able to synthesize other types of flavonoids. It has been suggested that the Caryophyllineae should be removed from Centrospermae on the basis of this single chemical difference only. Although a very large number of the Centrospermae has been tested, the majority still remain to be investigated. It is particularly interesting from a taxonomic point of view that betacyanins are common to Centrospermae and Cactales, orders which have long been considered to be phylogenetically related.

It is possible that the Centrospermae and Cactaceae have developed from primitive cormophytes producing nitrogenous colouring matter of betacyanin type and that they have retained this peculiarity throughout the ages because the betacyanins serve the same purpose as the anthocyanins and just as efficiently. The possibility that the Centrospermae is a biphyletic order cannot be excluded.

Unique or very rare compounds are often of little taxonomic interest but the situation may suddenly change as the result of unexpected discoveries. Ricinine (3) from castor bean, *Ricinus communis*, was long considered to be an odd compound of no taxonomic interest. However, recently a chemically related alkaloid, nudiflorine (4) was discovered in

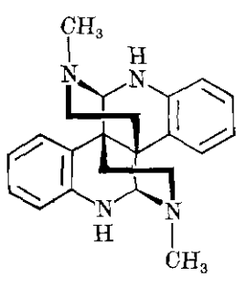


(3) Ricinine

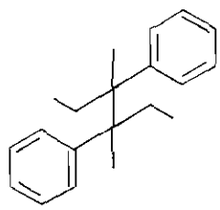


(4) Nudiflorine

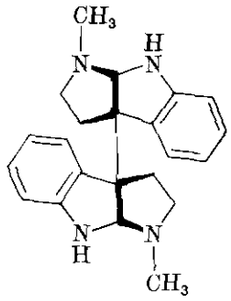
Trewia nudiflora. Both *Ricinus* and *Trewia* are monotypic genera and for that reason not very suitable for molecular taxonomic investigations. The genera are botanically closely related, both belonging to the tribe Acalypheae of the subfamily Crotonideae of Euphorbiaceae and this is obviously reflected in their alkaloids. There are several similar cases but it may suffice here to mention only calycanthine (5) and chimonanthine (6), chemically related alkaloids from the very closely related genera *Calycanthus* and *Chimonanthus* (both Calycanthaceae). One *Calycanthus* species in fact contains alkaloids of both types.



(5) Calycanthine



Carbon skeleton common to 5 and 6

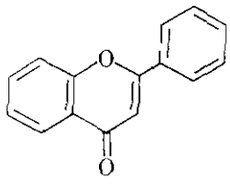


(6) Chimonanthine

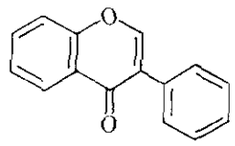
The occurrence of such closely related compounds in botanically related plants cannot be accidental. However, if by chance ricinine should be discovered in a grass this would be a curious coincidence but not a finding of any taxonomic value.

Chemical complexity and taxonomic value

Compounds of restricted distribution are taxonomically the most interesting. Simple flavones (7) are widely distributed but isoflavones (8) have been found only in a few plant families. Thus the isoflavones have greater taxonomic value than the flavones.

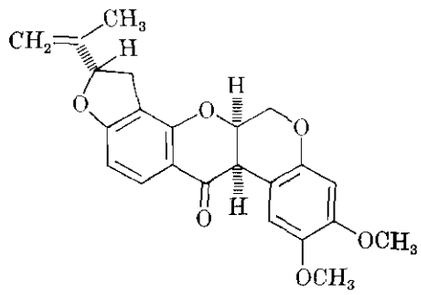


(7) Flavone

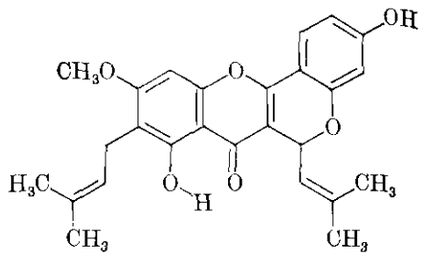


(8) Isoflavone

There are also many flavones which have become more or less modified by reductive, oxidative or substitution processes. The same is true of isoflavones, coumarins and many other simple structural types. The rotenoids (9) of Leguminosae, cycloartocarpin (10) from Moraceae and certain complex coumarins from Umbelliferae, Guttifereae and Rutaceae can serve as examples. Such "unusual" compounds have mostly a restricted distribution and therefore greater taxonomic interest than their simpler "parent compounds". The very complex steroids, however, are almost omnipresent. Obviously it is not the structural complexity as such that makes a compound valuable in



(9) Rotenone



(10) Cycloartocarpin

molecular taxonomy but the complexity and number of the biosynthetic pathways involved in its formation.

Patterns of constituents

In most comparative chemical investigations attention is paid very unfortunately only to a single group of compounds. Like their biological colleagues, chemists are often very specialized as regards interests and techniques. No taxonomist, however, would endeavour to classify a plant simply on the basis of a single characteristic. It has already been mentioned that certain compounds of restricted distribution do occur in widely different plants and that this has been used as an argument against molecular taxonomy. This is as unjustified as many similar proposals by over-optimistic chemists who on the basis of a single compound suggest that a genus should be split into two subgenera or that a species should be transferred from one genus to another.

Taxonomic conclusions of any value must be based on a large body of evidence and chemical characteristics are by no means superior to the classical characters. Moreover, the biological classification on the level of genera and subgenera is often subjective and sometimes simply serves practical purposes. It is useless to force doors open. A brief consultation with an experienced taxonomist, would, it is hoped, save chemists from such calamities.

Nicotine has been found in plants belonging to several families including Equisetaceae, Lycopodiaceae, Rosaceae, Leguminosae, Asclepiadaceae, Solanaceae and Compositae. This does not however mean either that nicotine is taxonomically useless or that these families are related. The important point is that in these cases nicotine is accompanied by quite different chemical constituents.

The heartwood of several *Sorbus* species contains unusual hydroxydiphenyl derivatives, the aucuparins, as well as a lignan xyloside

lyoniside. The latter compound occurs also in the leaves of an Ericaceae, *Lyonia ovalifolia* and aucuparin has been found in the wood of a Guttiferae, *Kielmeyera coriacea*. There is no lyoniside in the latter species but a series of xanthones common to many other species of Guttiferae. Many similar examples could be mentioned. It is clear that the co-occurrence of a compound and a simple derivative of it, e.g. a flavone and its methyl ether or glucoside, is taxonomically much less interesting than its co-occurrence with quite different types of compounds, e.g. a flavone and an alkaloid, a tropolone or an acetylene.

It sometimes happens that a compound which at first appears to be characteristic of a genus is lacking in one or more species. Such gaps can often be bridged with other types of compounds.

The value of the chemical approach to taxonomy is that it brings additional and independent evidence to a taxonomic discussion. However the botanical classification of plants, although not infallible, is the result of very careful considerations of classical characteristics and it is necessary to use equal care before drawing any taxonomic conclusions from chemical results. The chemist has therefore to investigate the widest possible range of compounds and their occurrence in several different organs of the group of plants concerned.

Chemically and biologically identical compounds

In molecular taxonomy the criterion of identity of chemical characteristics is different from that for chemical compounds. Two compounds from different species may have the same structure and configuration but they must be considered to be biologically different if they are produced along different biosynthetic routes. A single example will suffice to illustrate this point.

The amino acid lysine is formed in bacteria and blue-green algae from aspartic acid via α,ε-diaminopimelic acid but in yeast and *Neurospora* it is produced by a complex series of reactions involving the transamination of acetate-derived α-ketoadipic acid.

Clearly, the lysines from bacteria and yeast are biologically different compounds. Chemically identical compounds derived from these lysines are therefore biologically different. However, if the analogous lysines are coupled with biotin, forming biotin enzymes of identical catalytic properties, then the origin of the lysines is at the present stage a matter of little taxonomic interest since they are incorporated in the biochemical machinery, not in a secondary product.

It can be very difficult to say at what stage biosynthetic reactions begin to diverge from those of primary to those of secondary metabolism. The problem of biochemical identity or dissimilarity is still very

unclear and the elucidation of the biosynthesis of natural products is therefore a matter of fundamental importance in molecular taxonomy.

MOLECULAR TAXONOMIC INVESTIGATIONS

Most molecular taxonomic investigations originate and develop as follows. A novel type of organic compound is, rather accidentally, isolated from a natural source. It is often found to co-occur with compounds of similar chemical structure. They may differ as regards state of oxidation, number of hydroxyl groups or methoxyl groups but they all very obviously constitute a variation on a theme. This naturally arouses curiosity as to whether or not members of the new family of compounds also occur in related organisms. This is frequently found to be the case. One of the minor constituents in one species may be a major one in another. If the interest of the investigator is great enough, he will include more and more species in his research and perhaps ultimately cover the whole genus. If this is large and contains the same or closely related constituents throughout, then the results may be of very considerable taxonomic value. If the genus is large and pandemic the acquisition of starting material is difficult and the investigation is often discontinued in an incomplete state. It may be resumed later by workers in other parts of the world, to whom the remaining species are more easily accessible. This makes the screening of large taxa an irritatingly slow and haphazard process. The situation is familiar to biologists who often complain about the lack of modern, thorough, monographic treatments of entire genera or families.

At the present stage molecular taxonomy mainly involves the systematic screening of taxa for the occurrence of secondary metabolites and compilations of the distribution of compounds or types of compounds using some biological classification as a background. It thus differs little from phyto- or zoo-chemistry.

Much of this essential pioneering work requires a chemist but it will, no doubt, increasingly be taken over by biologists as analytical methods are improved and simplified. Close collaboration between biologists and chemists is needed for the sound development of molecular taxonomy. The chemist needs advice of biologists in choosing topics that are of the greatest taxonomic importance as well as assistance with the correct identification of biological starting materials.

Valuable chemical investigations have sometimes been published on compounds isolated from a plant identified simply by a local or trade name! Interestingly, on more than one occasion, phytochemists have been able to disclose erroneous identifications of plants on the ground that the reported constituents did not fit in well with phytochemical

experience—and have been able to suggest, correctly, which plant was actually investigated. Such incidents strengthen confidence in molecular taxonomy (cf. Hegnauer, 1967).

Small plants are usually extracted as such but of large plants frequently only a special organ is investigated. It therefore happens that much is known about the chemistry of the root or bark of one plant, the heartwood of a closely related species and the seeds of a third. It is necessary to keep this in mind when comparisons are made. In the wood of most species of the family Pinaceae resin ducts occur containing resin acids such as abietic acid. The wood of the true cedars (*Cedrus*) normally does not contain resin ducts and abietic acid is absent. Here the abietic acid is confined to the cones.

In this connexion it should perhaps be mentioned that we know almost nothing about the chemistry of the individual cells of an anatomically homogeneous tissue. Many compounds differing for example as regards state of oxidation may well be produced in different cells. Present-day cyto-chemists are usually more interested in molecular biology than in secondary metabolites.

Parasitism and symbiosis sometimes give rise to interesting metabolic changes. The ascomycete *Ceratostomella fimbriata*, when grown on *Ipomoea batatas*, produces a series of unusual isoprenoid compounds, among others a sesquiterpene, ipomoeamarone. Lichens almost regularly contain peculiar acids, both aliphatic and aromatic. They are sometimes produced in very large amounts and are useful for the identification of lichens. Compounds related to lichen acids have occasionally been found in fungi and it is generally believed that the fungal component of lichens is responsible for their formation. However, it is obvious that the symbiosis between fungus and alga greatly stimulates the synthesis of lichen acids. Certain orchids such as *Orchis* and *Laroglossum*, live in symbiosis with fungi (*Rhizoctonia*) and this symbiosis is regulated by the production of specific phenols of phenanthrene type (orchinol, hircinol). Wood of *Prunus jamasakura* contains flavonoids. When infested by *Polystictus* (*Coriolus*) *versicolor*, however, it contains large amounts of the lignan iso-olivil. This compound is foreign to both host and parasite but is a normal constituent of *Olea cunninghamia*. Infections are not always easily detected, at least not by chemists, and may cause grave mistakes.

AN EXERCISE IN MOLECULAR TAXONOMY. THE CONIFERS

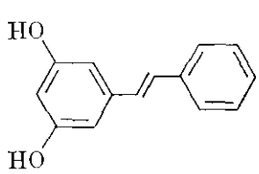
1. *Pinus*

I should like to illustrate a few of the above points by some work on the conifers which has partly been carried out by our group in

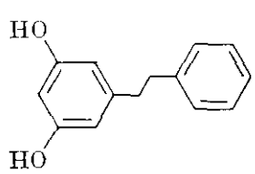
Stockholm. For various reasons the pines were the first to be investigated, and of the different organs the heartwood was chosen because it is dead and not subject to seasonal variation and other external influences.

The genus *Pinus*, embracing some 80–100 good species, has been carefully studied by several taxonomists. It is a very sharply defined genus and there is general agreement that it should be divided into two subgenera, *Haploxylon* and *Diploxylon*. The former pines have needles containing one vascular bundle, the latter have two. These subgenera have been further divided into sections and subsections. At this level, however, the various authors do not agree.

The first pine heartwood investigated by us was our common *Pinus silvestris*, Scots pine, a member of the subgenus *Diploxylon*. It contains a series of phenolic compounds of which only the stilbene derivative pinosylvin (11) and the dihydroflavones pinocembrin (13) and pinobanksin (14) will be mentioned here. Several hundred samples of Scots

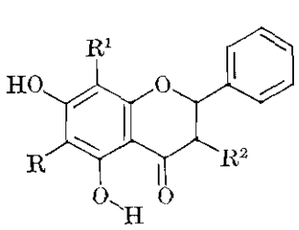


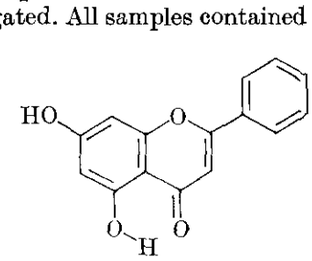
(11) Pinosylvin



(12) Dihydropinosylvin

pine were investigated (Erdtman *et al.*, 1951). Most of them were of Swedish origin but pine wood from Ireland, Spain and Poland and non-European plantations has also been investigated. All samples contained





(18) Chrysin

(13) ($R=R^1=R^2=H$): Pinocembrin
(14) ($R=R^1=H$; $R^2=OH$): Pinobanksin
(15) ($R^1=R^2=H$; $R=CH_3$): Strobopinin
(16) ($R=CH_3$; $R^1=R^2=H$): Cryptostrobin
(17) ($R=R^1=CH_3$; $R^2=H$): Demethoxymatteucinol

the above phenols. Some small quantitative differences could be noticed between pines from southern latitudes and from northern Sweden. This

is not surprising. Morphologically the pines vary considerably and several geographical "races" are known.

A few other pine species were then investigated and the results were so promising that it was decided to include as many species as possible. This work was undertaken by Dr. G. Lindstedt (cf. Lindstedt, 1951) and it was found that all diploxylon pines investigated (35 species) contained pinosylvin and the above dihydroflavones. Thus, from the point of view of heartwood phenolics the diploxylon pines all look very much alike.

In his revision of the genus *Pinus* Shaw (1914) recognized two sections and seven subsections in *Diploxylon* and two sections and six subsections in *Haploxylon*. Lindstedt also investigated fourteen haploxylon pines. The diploxylon phenols were present in practically all of them, but pinosylvin was accompanied by its hydrogenation product, dihydropinosylvin (12) and pinocembrin by its dehydrogenation product chrysin (18). There were also some interesting *C*-methylated flavonoids, e.g. strobopinin (15) and cryptostrobin (16) which occurred in species belonging to Shaw's section Cembra, subsection Strobi and section Paracembra, subsection Gerardianae.

If from these results we had endeavoured to suggest a classification of the pines (which, of course, we did not do) we would have separated the pine genus into two subgenera, identical with *Diploxylon* and *Haploxylon*. That corresponding to *Haploxylon* would have been further divided into two sections.

Chrysin and chrysin glycosides have been found in *Populus*, *Prunus*, *Pyrus*, *Docynopsis* and *Scutellaria*, pinocembrin and its glycosides in *Prunus*, *Sarothamnus*, *Aniba* and *Xanthorrhoea* and it would not be surprising if pinosylvin was found for example in a Leguminosae. The important thing is that all these substances co-occur in the pines. Together with the lignins of the pines, which are of the "conifer type" and differ from those of the angiosperms, the pine heartwood phenolics differentiate the pines from all other plants.

Nevertheless it seemed advisable to subject a biochemically entirely different class of pine constituents to a complementary comparative study. The most suitable group was the terpenoids. Much was already known about the diterpene acids ("resin acids") of the pines and the monoterpenes had been systematically studied by Mirov (1961). The dominating monoterpenes are the pinenes but *P. contorta* for example produces β-phellandrene. Crosses between this pine and the α-pinene-producing *P. banksiana* produce mixtures of these terpenes. *P. jeffreyi* which Shaw regarded as being a variety of *P. ponderosa* (*Diploxylon*, subsection Australes) produces a strange turpentine consisting almost entirely of n-heptane, a hydrocarbon which is in no way related to the

terpenes. In *P. ponderosa* Mirov found no heptane but a mixture of α- and β-pinene and Δ^3-carene. For this reason Mirov regards *P. jeffreyi* as a good species and since *P. sabiniana*, which was placed by Shaw in the subsection Macrocarpae is a great heptane producer, Mirov (1961) suggests that *P. jeffreyi* should be transferred to the Macrocarpae. *P. jeffreyi* hybridizes easily with *P. ponderosa* and the offspring contains constituents from both parent species. It would be a matter of great interest to investigate whether there are any differences between the resin acids of heptane and terpene producing pines. Several pines belonging to Shaw's Macrocarpae contain mixtures of heptane (and undecane) and terpenes and this is also true of several haploxylon pines. Possibly most pine turpentines contain at least traces of paraffin hydrocarbons. There are some recent indications that this may be the case (Smith, 1964). Mirov's work has shown that pines produce specific, constant, patterns of monoterpenes but they do not aid in distinguishing between the haploxylon and diploxylon pines.

During these important studies Mirov also paid some attention to the sesquiterpenes of the pine oleoresins. Before the gas chromatographic era this was an extremely difficult task. Nevertheless several important sesquiterpenes were characterized. Sesquiterpenes have more varying structures than monoterpenes and for that reason should be more interesting from a taxonomic point of view since one would expect that the formation of sesquiterpenes requires more complex enzyme systems.

Dr. Westfelt (1967) has recently subjected the turpentine fractions of *Pinus silvestris* to an extensive investigation. Prior to his work only a few neutral sesquiterpenes were known from this species and only a single neutral diterpene. Westfelt isolated 15 sesquiterpenes and 13 diterpenes. This shows not only the efficiency of modern methods but also the amount of chemical characters of terpene type we can expect to know when the whole genus has been covered. With present day analytical methods this can be done fairly easily.

We were, of course, anxious to see if there was any indication of differences between the higher terpenes in *Diploxylon* and *Haploxylon* and therefore Westfelt investigated also the neutral sesqui- and diterpene fractions of one haploxylon pine, *P. albicaulis*. The sesquiterpene fraction was simpler than that of *P. silvestris*. The main diterpene hydrocarbon was the macrocyclic thunbergene (Mirov's cembrene) which perhaps is a characteristic constituent of at least some groups of pines.

Much work is now being done in different laboratories on the constituents of pine barks, needles and seeds and the results have been very interesting. When the distribution of the new chemical pine characteristics is known we shall possess a chemical map of the genus which

together with biological evidence may enable us to produce an improved classification of the pines (Mirov, 1967).

2. *Pinus* or *Ducampopinus*?

After having gained some experience in the chemistry of the pines we were anxious to investigate that of *Pinus krempfii*. This is a most unusual pine, possessing broad leaves in pairs. The leaves have a single vascular bundle and the tree consequently should be a haploxylon pine. However, it is so aberrant that some taxonomists have raised it to the rank of a subgenus or genus of its own, "*Ducampopinus krempfii*". Some authors apparently have even considered it to be a hybrid of a pine and *Keteleeria* or *Pseudolarix* (Dallimore *et al.*, 1966). *P. krempfii* grows in the Nhatrang region of Viet-Nam and under the prevailing circumstances it was not easy to obtain a wood sample. However, a small piece containing sapwood as well as heartwood was ultimately secured. The amount was sufficient only for an examination of the phenolics and the resin acids. The latter were those commonly encountered in pine heartwoods and the phenolics were typical of *Haploxylon* (Erdtman *et al.*, 1966b). There was only one new compound, demethoxymatteucinol (17) formerly isolated from a fern (*Matteucia*).

The pattern of phenolics isolated from *P. krempfii* was so similar to that of Shaw's subsection Gerardianae (and some of the Strobi) that a reinvestigation of some of these species was undertaken (Erdtman *et al.*, 1966a). It turned out that both *P. bungeana* and *P. gerardiana* (Gerardianae) as well as *P. peuce* (Strobi) contained demethoxymatteucinol. Hence from a chemical point of view *P. krempfii* undoubtedly falls within the subgenus *Haploxylon*. Taxonomically this is, of course, a very minor result but it shows the value of chemical characteristics when there is sufficient background information.

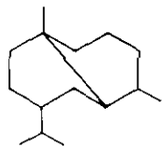
I have dealt with pine chemistry at some length because it shows:

1. That the whole genus possesses a characteristic heartwood chemistry;
2. That there are clear-cut chemical differences between the botanical subgenera;
3. That all the heartwood constituents taken together characterize a pine;
4. That there are indications that in future one may be able to characterize categories of even lower rank than the subgenera, particularly when more is known about the chemistry of other organs of the pines.

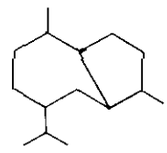
These results will probably not impress a botanical taxonomist and if he concludes that nothing new of taxonomic interest has come out of this work I am quite ready to agree. Our investigation should be

regarded primarily as an examination of the scope of molecular taxonomy rather than as a practical application.

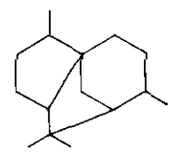
Taxonomists (or systematists) would be much more interested to



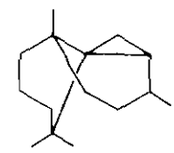
(19) Selinane



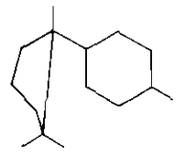
(20) Guaiane



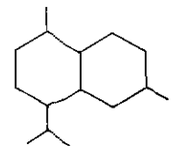
(21) Cedrane



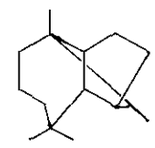
(22) Thujopsane



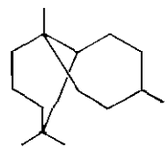
(23) Cuparane



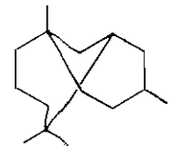
(24) Cadinane



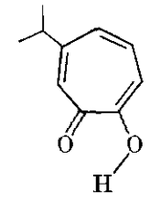
(25) Longifolane



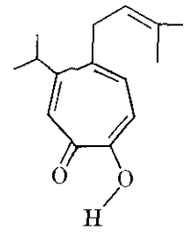
(26) Himachalane



(27) Allohimachalane



(28) β-Thujaplicin



(29) Nootkatin

learn if it is possible to say anything about the interrelations of the genera of the family Pinaceae on the basis of their chemical constituents. Based on morphological and anatomical evidence, botanists have recog-

nized Pinaceae as being a natural group. The other genera of Pinaceae have not yet been nearly as extensively explored as *Pinus*, but there is nevertheless much chemical evidence (flavonoids, lignans, resin acids) showing that this family is, indeed, a natural one differing distinctly from other conifer families (Erdtman, 1952, 1956, 1963).

3. Cupressales

There are very great chemical differences between the orders Pinales and Cupressales—and some similarities. The chemistry of Cupressales has recently been reviewed (Erdtman and Norin, 1966) and abounds in interesting features. There are several types of sesquiterpenes which appear to be limited to Cupressales such as those belonging to the selinane (19), guaiane (20), cedrane (21), thujopsane (22) and cuparane (23) types.

Common to Pinales and Cupressales are sesquiterpenes with cadinane (24) and longifolane (25) skeletons. Unique, so far, for *Cedrus* are sesquiterpenes of himachalane (26) and allohimachalane (27) type.

Toxic tropolones of modified terpene type such as β-thujaplicin (28) and nootkatin (29) have only been found in the Coniferae and then only in certain genera of the family Cupressaceae, never in any of the heterogeneous family Taxodiaceae. There are, however, several other compounds which one would call typical of the Cupressaceae which do occur in Taxodiaceae, e.g. cedrol in *Sciadopitys* and *Cunninghamia*. The more complex tropolone nootkatin seems to be restricted to *Cupressus* and *Juniperus*, the sole exception known being the "Alaska yellow cedar", *Chamaecyparis nootkatensis*, a conifer of somewhat uncertain systematic position that was formerly regarded as constituting a monotypic genus and called *Callitropsis nootkatensis*. There are several chemical differences between *Ch. nootkatensis* and other *Chamaecyparis* species and it might be worth while undertaking a botanical reconsideration of the taxonomic position of *Ch. nootkatensis*.

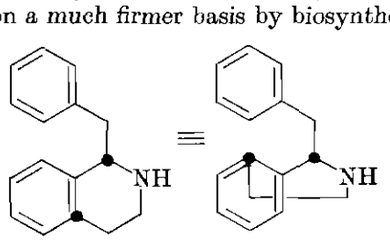
The presence or absence of tropolones is an important molecular taxonomic character in the Cupressaceae even though tropolones are absent from some species of genera where tropolones are to be expected. An example is *Juniperus virginiana*. However here the terpenoid constituents, which are closely similar to those of the other juniper species that have been investigated, serve to bridge the tropolone gap. The reason for these occasional anomalies is unknown but there is of course the possibility that the tropolone content is so low that none can be discovered even with our most powerful present day methods. The conifer tropolones are undoubtedly biochemically related to the terpenes. Presumably relatively small mutations can block the chain of reactions leading to the tropolones.

BIOSYNTHESIS

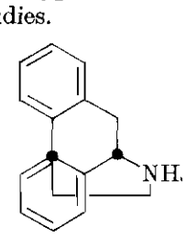
Biogenetic hypotheses

It has already been mentioned several times that a knowledge of the biosynthesis of secondary products is essential for the development of molecular taxonomy. Early attempts by organic chemists to throw some light on the biosynthetic processes led to the "biogenetic" or "intuitive" hypotheses proposed by Winterstein, Robinson, Ruzicka and many others. They were based upon comparisons of the molecular architecture of natural compounds and intelligent "dissections" of their structures into "building stones" of biologically acceptable nature. They also included some taxonomic and biochemical elements. It was never claimed, however, that they were anything more than indications of general trends along which Nature probably works.

Although highly speculative and, naturally, received with great scepticism by biochemists, the biogenetic hypotheses had and still have an enormous impact on structural organic chemistry. They could be used so successfully in very different domains of natural product chemistry that there was little doubt that many of them contained fundamental principles in biosynthesis. They could frequently be used for sorting out the most probable of several possible structures; Robinson's interpretation of the structure of morphine is the classical example. Several structures for this alkaloid had been proposed but none appeared to be quite acceptable. Robinson suggested the correct structure because it could be dissected into two identical halves derived from phenylalanine molecules and combined as in several papaveraceous benzylisoquinolines (30–31). Some of these hypotheses have been put on a much firmer basis by biosynthetic studies.



(30) Identical benzyl isoquinoline structures



(31) Basic skeleton of morphine

General pathways in the biosynthesis of secondary metabolites

1. FATTY ACID PATHWAY

As the result of brilliant investigations by several groups of biochemists many aspects of primary metabolism have been elucidated and

some general routes of particular interest in relation to the biosynthesis of secondary products emerged. They include the fatty acid route leading from acetate-malonate to fatty acids. Lynen has shown that the fatty acid synthesis takes place in highly organized multienzyme complexes, so-called fatty acid synthetases, as illustrated very schematically below.

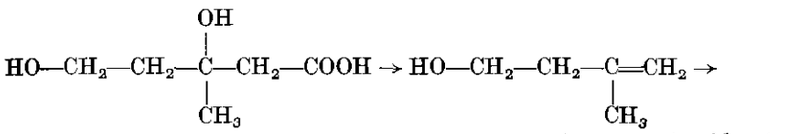
$$CH_3—COOH \xrightarrow{CH_2(COOH)_2} CH_3—CO—CH_2—COOH \rightarrow \rightarrow$$
$$CH_3—CH_2—CH_2—COOH \xrightarrow{CH_2(COOH)_2}$$
$$CH_3—CH_2—CH_2—CO—CH_2—COOH \rightarrow \rightarrow$$
$$CH_3—CH_2—CH_2—CH_2—CH_2—COOH \xrightarrow{CH_2(COOH)_2}$$

etc. etc.

The "fatty acid pathway" appears to be very general and leads directly to fatty acids which would not be characterized as secondary products although they are often accumulated or stored. Secondary transformations of the primary fatty acids sometimes give rise to the formation of compounds of taxonomic interest, e.g. unsaturated fatty acids of unusual types, acetylene derivatives, hydroxy acids etc.

2. MEVALONIC ACID PATHWAY

Another very general route is the mevalonic acid route leading from acetic acid via mevalonic acid (32) to the isopentenyl units (33) which are the building stones of terpenoids, steroids, rubber etc.



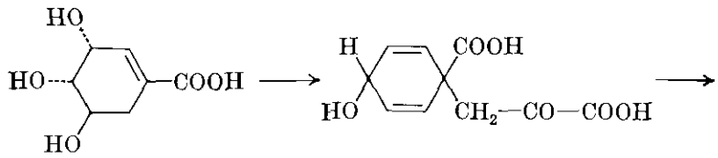
(32) Mevalonic acid (33) Isopentenylalcohol

3. SHIKIMIC-PREPHENIC ACID PATHWAY

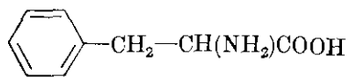
The shikimic–prephenic acid pathway leads from carbohydrates via shikimic acid (34) to a large variety of aromatic compounds, e.g. phenylalanine (35) and lignins.

4. POLYKETIDE PATHWAY

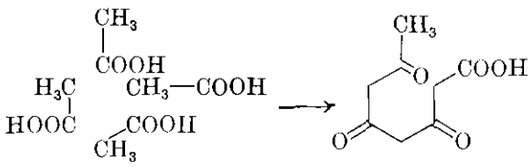
There is one general route, leading from acetic and malonic acids (or an acid: R—COOH and malonic acid) to a very large number of natural products, which has essentially been inferred from comparisons of the structure of naturally occurring substances, the so-called polyketide pathway. The hypothesis originated from Collie's early speculation on the role of acetic acid in biosynthesis, was further developed by Robinson and was much used by him in his lectures as early as 1930 (as far as



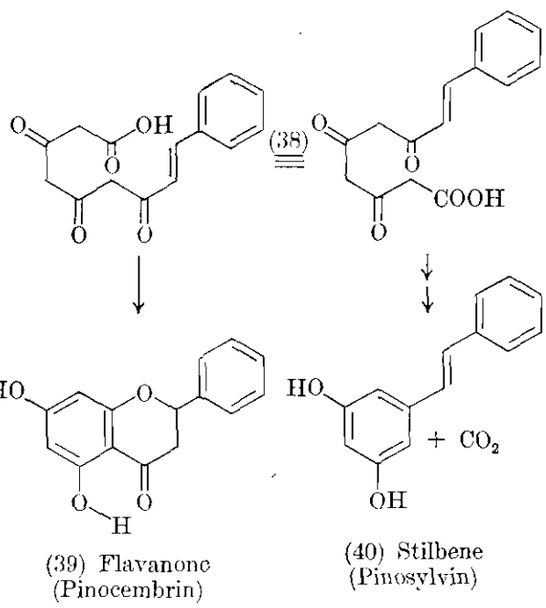
(34) Shikimic acid Prephenic acid



(35) Phenylalanine



Four molecules of acetic acid (36) (37) Orcellinic acid



(39) Flavanone (Pinocembrin) (40) Stilbene (Pinosylvin)

the author's experience goes). It was greatly extended and generalized by Birch (1957). The polyketide hypothesis is still to a large extent

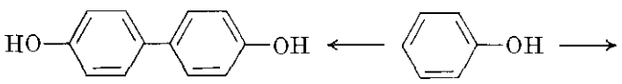
"biogenetic" but has received much support from biosynthetic experiments by many chemists including Birch himself. The essence of the polyketide hypothesis is best illustrated by some simplified examples (in which the role of malonic acid is left out).

Four molecules of "acetic acid" are condensed to form orcellinic acid (36–37).

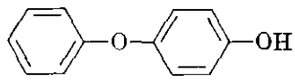
In an analogous manner a molecule of another acid, for example cinnamic acid, is condensed with three molecules of acetic acid (38). Cyclization of 38 can take place in two different ways, leading (a) to a flavonoid (39) and (b) to a stilbene (40).

It appears that the polyketide and the fatty acid routes are related and Lynen (1967) tentatively suggests that the polyketide syntheses take place in multi-enzyme complexes.

There are, of course, many additional enzymatic systems operating which produce variations of the main themes, condensations, reductions, oxidations, hydroxylations etc. Among such modifying factors are the enzymes that give rise to the oxidative coupling of phenols (e.g. 41) leading to diphenyl (42) or diphenyloxide (43) derivatives. Such reactions presumably involve the intermediate formation of free radicals (Barton and Cohen, 1957; Erdtman and Wachtmeister, 1957).



(42) (41) Phenol



(43)

Some examples of biosyntheses of secondary metabolites

Most of the biochemical work on primary metabolism has been done using mutants or radioactively labelled compounds but as soon as the latter became available many organic chemists began to test the biogenetic hypotheses for the biosynthesis of secondary products. Sometimes these hypotheses were confirmed, at least in principle, sometimes they were rejected. The methods used are simple in theory. Specifically labelled compounds presumed to function as precursors in the biosynthesis of a natural product are administered to an organism known to produce it. After some time the product is isolated. If labelled the compound is systematically degraded in such a way as to permit the examination of all its component atoms separately. (At least in the ideal

case.) In this way it is possible to locate the label derived from the introduced and incorporated "precursor". Such investigations, simple as they look on paper, require refined experimentation. Many very elegant studies of this type have now been carried out.

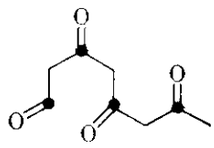
The results must be considered with great care. Failure to accept an introduced substance does not necessarily mean that it cannot serve as a precursor. Racial and seasonal variations, external influences such as light and temperature may cause complications. Perhaps the greatest trouble is the low incorporation which is often observed, frequently amounting only to a small fraction of a per cent. This sometimes makes one uncertain whether the product isolated has been synthesized along a main route or comes from some minor, alternative pathway. Low incorporation is often ascribed to slow diffusion of the labelled compound into the active centres of the cell. The situation is much simpler if it can be introduced in a biologically more normal way as when microorganisms are "fed" or animal organs are perfused via the blood stream. Often, however, an excised leaf or stem is simply dipped into a solution of the labelled compound or the compound is injected into a tissue or below the bark. In such cases one cannot exclude the possibility that the compound is largely taken up by cells not normally involved in the reaction studied. What they will produce when compounds strange to them are forced upon them, and how the reaction products will affect the normal metabolism, is unknown.

Sometimes the introduced substance is first biologically degraded into simple molecules which are then used for syntheses. In such cases the isolated secondary product will be more or less randomly labelled and this is mostly easily detected. If compounds carrying several suitably located labels are used it is possible to find out whether the whole molecule or only part of it is incorporated. Sometimes several different molecules can be accepted and transformed into the same end-product. This may indicate that there are several biosynthetic paths to the same product. Mostly such molecules are chemically closely related and are considered to be "biochemically equivalent". It is important to remember that the successful incorporation of a molecule does not necessarily mean that it constitutes a natural intermediate in a chain of biological syntheses.

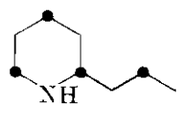
1. CONIINE

From a biogenetic point of view it appeared possible that coniine (45) is derived from lysine and some early experiments indicated that this could be correct. However, feeding *Conium maculatum* with CH_3—$^{14}COOH$ gave rise to coniine labelled as indicated by dots in the formula (45) and it is now believed that a polyketide formed from

four molecules of acetic acid (44) is involved in the synthesis. Apparently many other alkaloids have a similar origin. There are also alkaloids which must be derived from mevalonic acid.



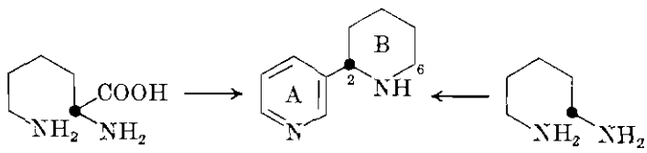
(44)



(45) Coniine

2. ANABASINE

Nicotiana glauca is a plant which appears to form lysine from 2,6-diaminopimelic acid. It produces anabasine (47), an alkaloid structurally resembling nicotine. It has been suggested that the piperidine ring (B) of anabasine is derived from lysine. (The biosynthesis of the pyridine ring (A) will not be discussed here.) Feeding *Nicotiana glauca* with lysine labelled as shown in formula 46 gave anabasine labelled as in formula 47.



(46) Lysine (47) Anabasine (48) Cadaverine

It follows that it is the carbon atom next to the carbonyl group which becomes attached to the pyridine ring. Cadaverine, a decarboxylation product of lysine, singly marked as in formula 48 was an "excellent precursor" to the B-ring. The resulting anabasine, however, was equally labelled at C(2) and C(6) and it can be concluded that the route from lysine to anabasine does not proceed via the symmetrical "lysine equivalent" cadaverine.

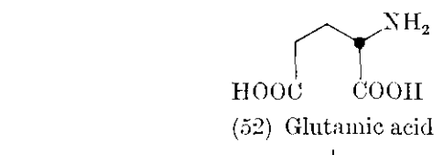
3. NICOTINE

Most other *Nicotiana* species produce nicotine but not anabasine. By analogy with the biosynthetic route postulated for anabasine it would be expected that the pyrrolidine ring of nicotine is derived from the next lower homologue of lysine, ornithine, and ornithine is in fact incorporated in feeding experiments. There is one unexpected feature, however. Ornithine labelled as in formula 49 gives nicotine (50) labelled in position 2 as well as in position 5 and this requires that a symmetrical

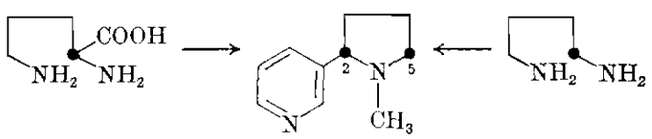
entity is interposed between ornithine and nicotine. This can be explained in terms of mechanistic organic chemistry—which are supposed to be applicable also in biological reactions.

"Biological equivalents" of ornithine, e.g. putrescine (51), proline and glutamic acid (52) are incorporated into nicotine in the same way as ornithine.

These results look straightforward and most organic chemists have accepted them as providing answers to the biosynthetic problem under discussion.



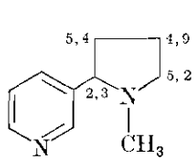
(52) Glutamic acid



(49) Ornithine (50) Nicotine (51) Putrescine

There is one snag, however, and that is the feeding technique itself. Under natural conditions plants are not "fed" and it is not certain that the alkaloid producing cells are normally provided with prefabricated amino acids synthesized in some more or less distant cells. In order to arrive at a true and detailed picture of the alkaloid synthesis it would be necessary to follow the example of Calvin during his brilliant researches on the path of carbon in photosynthesis.

The path of carbon in nicotine synthesis has recently been investigated by Rapoport (Liebman *et al.*, 1967). Short time exposure of plants to $^{14}CO_2$ gave nicotine whose pyrrolidine ring was labelled in a manner which greatly differed from that obtained in feeding experiments (53).



(53) Distribution of ^{14}C-labelling of the pyrrolidine ring of nicotine after short time feeding with $^{14}CO_2$

This uneven labelling pattern is not in agreement with the assumption of a symmetrical intermediate. Rather it indicates that the nicotine is formed in two or more different ways, the observed uneven labelling being due to a summation of the specific labellings.

Alternatively a precursor of nicotine, e.g. glutamic acid (52), is biosynthesized in *Nicotiana* by a novel route. Some very recent results of an investigation of the biosynthesis of glutamic acid in *Nicotiana rustica* using $^{14}CO_2$ seem to indicate that this may be the case (Burns *et al.*, 1967). Whatever the ultimate solution of this confused problem may be it clearly demonstrates the care required in drawing conclusions from feeding experiments using postulated precursors. It will now be necessary to check many biosynthetic theories using labelled carbon dioxide.

Anabasine is structurally related to nicotine and occurs in a few *Nicotiana* species (Solanaceae, Tubiflorae) sometimes together with nicotine. The possibility has not been ignored that the *N*-methyl groups of nicotine could be incorporated into the piperidine ring of anabasine by some biochemical reaction, for example an oxidative process, but there is no biosynthetic evidence for this.

Anabasine is a constituent of *Anabasis aphylla* (Chenopodiaceae, Centrospermae) where it co-occurs with alkaloids otherwise found in some Leguminosae. Clearly a detailed knowledge of the biosynthesis of nicotine and anabasine in unrelated plants would be of great general taxonomic value.

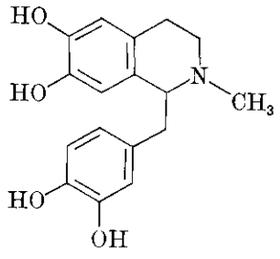
4. MORPHINE

Early *in vitro* syntheses modelled to simulate probable biosynthetic reactions in this field were carried out by Robinson and Sugasawa in 1932. They oxidized laudanosoline (54) in the hope of obtaining the aporphine, shown in formula 56. However, the reaction took another, *a priori* equally probable, route and the substance shown in formula 55 was formed. Interestingly, compounds of this type have since been found in nature.

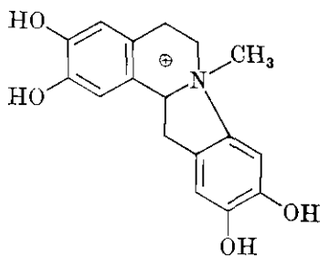
Among the most beautiful biosynthetic investigations which have been made using feedings with postulated precursors are those by Barton and Battersby and their groups on the opium alkaloids (Barton, Battersby *et al.*, 1965).

Laudanosoline was found to be incorporated into morphine. It seems first to be methylated to reticuline (57) which is then oxidized to the dienone salutaridine (58). Reduction and elimination of water follows (59), leading to thebaine (60) which is finally demethylated and reduced via codeine to morphine (61).

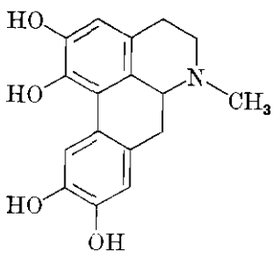
The various steps in the conversion of amino acids such as phenyl



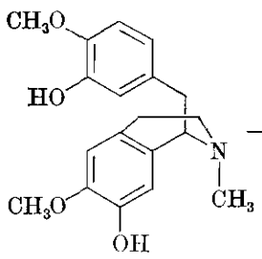
(54) Laudanosoline



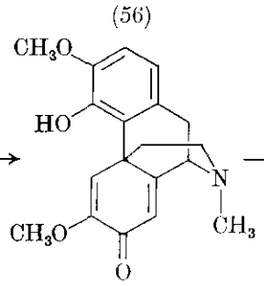
(55)



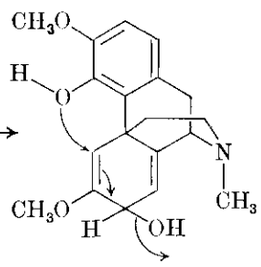
(56)



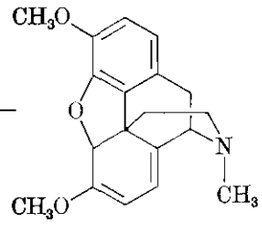
(57) Reticuline



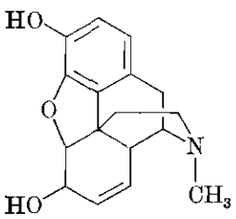
(58) Salutaridine



(59)



(60) Thebaine



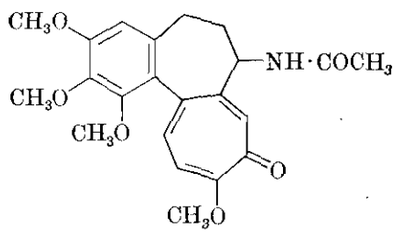
(61) Morphine

alanine and tyrosine to morphine are so well documented using labelled and multilabelled precursors that it seems probable that the natural biosynthesis essentially proceeds as indicated. Perhaps in the cell some of the intermediates never become liberated but are further transformed *in situ*. It is hard to believe that all these reactions take place in a single multi-enzyme complex.

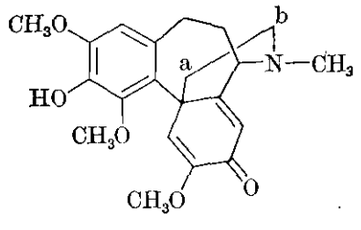
Reticuline was first discovered in *Annona reticulata* (Annonaceae) and salutaridine in *Croton salutaris* (Euphorbiaceae), plants of widely different systematic positions. It is obvious that these plants and *Papaver somniferum* have some alkaloid constructing enzyme systems which are closely analogous or perhaps even homologous. Interestingly, it has recently been found that reticuline as well as salutaridine also occur in the poppy.

5. COLCHICINE

The *Colchicum* alkaloid colchicine (62) is most unusual and its structure long remained a mystery. Fair guesses were that colchicine is somewhat related to the benzylisoquinolines and that an oxidative phenol coupling takes place at some stage during the biosynthesis. The compound contains a tropolone ring and it has been suggested that this is formed from a benzene ring by addition of a C_1-unit followed by ring expansion. However, it appeared likely that the additional carbon atom originated from a side chain. Just as the missing link *Archaeopteryx* was needed to give a clearer understanding of the relations between reptiles and birds the compound androcymbine (63) from *Androcymbium melanthioides*, a close relative of *Colchicum*, was needed before the colchicine structure could be properly appreciated.



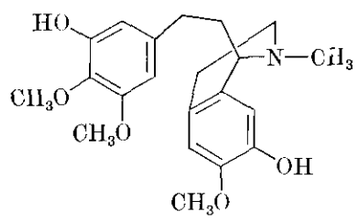
(62) Colchicine



(63) Androcymbine

When fed to *Colchicum autumnale* androcymbine methyl ether was transformed into colchicine in the extraordinarily high yield of 15% (Battersby, 1967). The bond between the carbon atoms a and b is cleaved and a takes part in the formation of the tropolone ring. The carbon atom b is eliminated as is the CH_3 of the *N*-methyl group. In these extremely beautiful studies the multilabelled compound shown in

formula 64 was synthesized and its biological transformation to colchicine (10% yield) was carefully investigated. It appears that



(64)

Colchicum and *Androcymbium* have several enzymes in common since *Androcymbium* also produces colchicine. Androcymbine is apparently not entirely strange to *Colchicum* and is therefore willingly accepted and transformed into colchicine.

Examples of compounds believed to be produced by several pathways

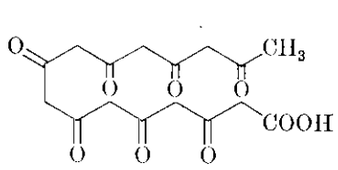
1. ANTHRAQUINONES

Anthraquinones are found both in animals (crinoids and insects) and in numerous systematically unrelated plants, and their taxonomic value has been much discussed.

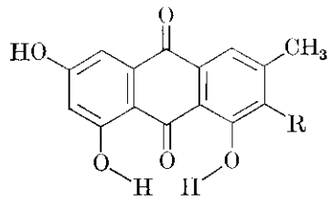
Endocrocin (66) occurs in some lichens and fungi and there is strong biosynthetic evidence that it is of polyketide origin (65).

Its decarboxylation product, emodin (67) and various analogues containing CH_2OH, CHO or COOH instead of the methyl group and methyl ethers have also been found in lichens and fungi. Emodin is a constituent of many angiosperms: *Rheum*, *Rumex*, *Polygonum* (Polygonaceae), *Rhamnus* (Rhamnaceae) and *Sonneratia* (Sonneratiaceae). Anthraquinones with a variety of hydroxylation patterns are known from several plant orders: Myrtales, Liliaeflorae, Polygonales, Rhamnales, Rubiales, Rosales and Tubiflorae. Formally they can all be derived from the hypothetic polyketide precursor (65) by simple oxidative and reductive changes and decarboxylations. As a rule the hydroxylation patterns of the fungal anthraquinones fit in better with the simple polyketide hypothesis than those of the anthraquinones from angiosperms—anthraquinone itself is known from *Schinopsis* (Quebracho, Anacardiaceae). It has therefore long been suspected that there is more than one route to the anthraquinones.

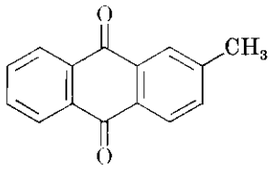
Teak wood (*Tectona grandis*, Verbenaceae) contains the anthraquinone tectoquinone (68) as well as the naphthoquinones lapachol and deoxylapachol (69) and the structural relations of these co-occurring compounds are immediately evident.



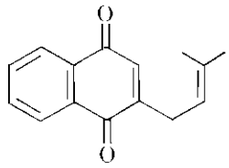
(65) Polyketide



(66) (R=COOH). Endocrocin
(67) (R=H). Emodin



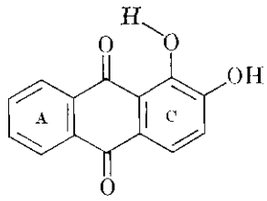
(68) Tectoquinone



(69) Deoxylapachol

The lapachols are isopentenyl derivatives of naphthoquinones; very probably the isopentenyl groups are of mevalonic acid origin. (There is, of course, a theoretical possibility that lapachol is formed by a cleavage of the tectoquinone molecule but this is most improbable.) Because of the occurrence of anthraquinones in crinoids it is of interest that several echinoderms produce a large variety of naphthoquinones.

Anthraquinones are common in Rubiaceae, alizarin (70) from *Rubia*



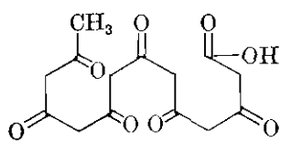
(70) Alizarin

tinctorum being the classical example. Leistner and Zenk (1967) have shown that acetate can be incorporated into ring C but that ring A is probably of shikimic acid origin. This is a further indication that anthraquinones can be produced in different ways in different plants.

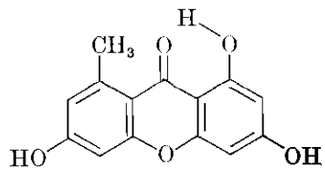
2. XANTHONES

Like the anthraquinones, xanthones have been found in micro-organisms as well as in higher plants (e.g. Anacardiaceae, Guttiferae, Gentianaceae). Formally, lichexanthone (72) from *Parmelia formosana*

can be derived from a single polyketide (71) containing seven "acetate units" or one unit less than that proposed for the endocrocin precursor.

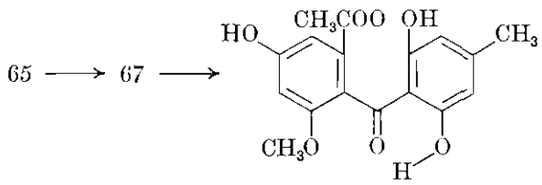


(71) Polyketide



(72) Lichexanthone

It has been believed that some xanthones may be derived from anthraquinones by oxidative fission of the central ring. Benzophenone carboxylic acids would then constitute intermediates. The benzophenone carboxylic acid ester sulochrin (73) from *Aspergillus terreis*

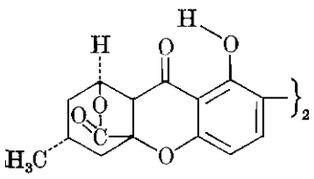


(73) Sulochrin

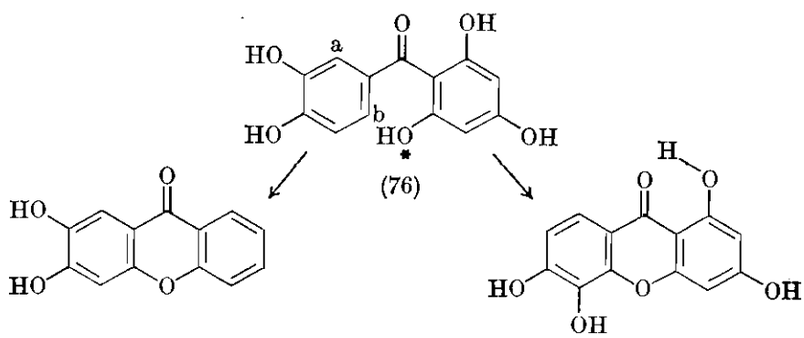
could be derived from emodin by an oxidative fission of the central ring *and thus from a single polyketide chain. However, biosynthetic experiments* using mutants indicate that, in fact, sulochrin is derived from two separate, acetate-derived polyketide precursors (Curtis *et al.*, 1966).

Ergoflavine co-occurs with endocrocin in ergot. It is a symmetrical double xanthone (74) in which the carboxyl group involved in the lactone ring may originally have been part of the central ring of an anthraquinone. The structures of some other ergot constituents are also consistent with formation by oxidative fission of anthraquinones.

Xanthones are common in the angiosperm family Guttiferae and for example the xanthones (75 and 77) co-occur in the heartwood of



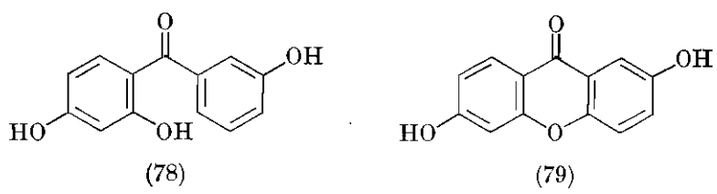
(74) Ergoflavine



Symphonia globulifera. It is believed that they are formed from benzophenone precursors by oxidative coupling of the phenol–diphenyl ether type. The *Symphonia* xanthones can be accounted for as oxidation products of the benzophenone (76) in which ring A is presumably of shikimic acid origin and ring B of acetate origin. Coupling can take place in two ways either * to *b*, giving (75) or * to *a*, giving (77).

There are several examples of similar relations between benzophenones and xanthones, e.g. in *Gentiana*, *Kielmeyera*, *Mesua* and *Maclura* and moreover, the benzophenone (78) has been oxidized *in vitro* to the xanthone (79) in high yield.

Many examples of possible or probable different roads to chemically identical or related compounds from other groups of natural products could be mentioned but a mere enumeration of the various hypotheses proposed would serve no purpose. What is needed are experimental studies of the actual biosyntheses. The results will be of great interest from a molecular taxonomic point of view.



HIGH MOLECULAR WEIGHT CONSTITUENTS

The biosynthesis of natural products is the result of chemical reactions catalysed by specific proteins, enzymes. These are sometimes organized as multi-enzyme complexes and may be localized in intracellular organelles. It is only recently that methods have been elaborated

for the separation and isolation of individual proteins and for the elucidation of their structures and molecular morphology, but progress in this field is now rapid and it can be expected that in a decade or so the number of structurally known proteins of similar biological function from species of different taxa will have increased to such an extent that detailed comparisons of extensive material will be feasible. Undoubtedly, some types of proteins will prove more suitable for taxonomic purposes than others. It is already known that in some proteins there are regions in which the amino acid patterns can vary within certain limits. Other regions are identical and essential for the tertiary structure of the proteins as well as for their enzymatic activities. The more variable sections in protein molecules appear to be particularly interesting in relation to taxonomy since their differences reflect mutational changes.

Serological techniques, which are based on specific properties, particularly of proteins, have been much employed for taxonomic purposes, in the beginning over-optimistically, as is often the case. However, great strides have been made in this field and the techniques have become highly refined. Although the results cannot yet be interpreted on a molecular basis serological results no doubt have taxonomic significance, just as spots and peaks of unknown compounds obtained in chromatographic studies can be useful, for instance for the characterization of hybrids. However, knowledge of the exact nature of chemical characteristics is always to be preferred. Cheap and rapid chromatographic screening methods are much used in modern taxonomy but fortunately the tendency is to favour more exact investigations.

The biosynthesis of the enzymes and their organization in the cells is governed by the DNAs from whose structural elucidation much information of taxonomic importance is expected. The supreme hierarchical position of the DNAs is, however, much debated. The inclusion of the high molecular weight substances in the description of species increases the basis for comparative studies more quantitatively than qualitatively. It is clear that a greater number of characteristics obtained from researches utilizing quite different techniques should enable systematists to draw increasingly well founded conclusions as to the interrelations of species.

It is still, however, beyond our ability to describe the dynamics of the evolutionary processes in terms of physics and chemistry, but scientists have not failed to speculate on the possibility of solving such problems by comparative studies of amino acid sequences in proteins and, for example, it has been suggested that it should be possible to "reconstruct" the amino acid sequences of ancestral proteins. Pending such dramatic events it may be advisable to focus attention on the

chemistry of hybridization and allopolyploidy, factors controlling hormone balance and morphogenesis etc.

Life can accomplish the synthesis of an enormous variety of molecules but its synthetic achievements are limited by the laws of physics and chemistry. If life is truly polyphyletic, i.e. has originated from several primitive beings independently formed from non-living matter, differences in the chemistry of the earliest organisms may have determined their further evolution along separate lines. A chemical theory of evolution is needed but there is as yet little experimental basis for such a theory. It is possible that the chemistry of speciation can be described but no chemist, I presume, would deny that selection has the final word in evolution.

GENERAL REFERENCES

1. *Molecular Taxonomy*

Alston, R. E. and Turner, B. L. (1963). "Biochemical Systematics." Prentice Hall, New York.

Hegnauer, R. (1962–5). "Chemotaxonomie der Pflanzen" I–IV. Birkhäuser, Basel.

Swain, T. (ed.) (1963). "Chemical Plant Taxonomy." Academic Press, London.

Swain, T. (ed.) (1966). "Comparative Phytochemistry." Academic Press, London.

2. *Biosynthesis*

Bernfeld, P. (ed.) (1963). "Biogenesis of Natural Compounds." Pergamon Press, Oxford.

Bu'Lock, J. D. (1965). "The Biosynthesis of Natural Products. An Introduction to Secondary Metabolism." McGraw-Hill, New York.

Richards, J. H. and Hendrickson, J. B. (1964). "The Biosynthesis of Steroids, Terpenes and Acetogenins." Benjamin, New York.

REFERENCES

Battersby, A. R. (1967). Biosynthesis of the indole and colchicum alkaloids. *Pure appl. Chem.* **14** (1), 117–136.

Barton, D. H. R. and Cohen, T. (1957). Some biogenetic aspects of phenol oxidation. "Festschrift A. Stoll." 117–43. Birkhäuser, Basel.

Barton, D. H. R., Kirby, G. W., Steglich, W. and Thomas, G. M.; and Battersby, A. R., Dobson, T. A. and Ramuz, H. (1965). Investigations on the biosynthesis of morphine alkaloids. *J. chem. Soc.* 2423–38.

Birch, A. J. (1957). Biosynthetic relations of some natural phenolic and enolic compounds. *In* Zechmeister, L. (ed.), *Fortschr. Chem. org. NatStoffe* **14**, 186–216.

Burns, L., O'Neal, R. and Koeppe, R. (1967). Labelling patterns in glutamic acid in *Nicotiana rustica* L. from carbon-14-dioxide. *J. Am. chem. Soc.* **89**, 3938–9.

Curtis, R. F., Harris, P. C., Hassal, C. H., Levi, J. D. and Phillips, D. M. (1966). The biosynthesis of phenols. Mutation and radioactive tracer studies regarding the biosynthesis of sulochrin. *J. chem. Soc.* (C), 168–74.

Dallimore, W., Jackson, B. and Harrison, S. G. (1966). "A Handbook of Coniferae and Ginkgoaceae." 438. Edward Arnold, London.

Erdtman, H. (1952). Chemistry of some heartwood constituents of conifers and their physiological and taxonomic significance. *In* Cook, J. (ed.), "Progress in Organic Chemistry. I". 22–63.

Erdtman, H. (1956). Organic chemistry and conifer taxonomy. *In* Todd, A. (ed.), "Perspectives in Organic Chemistry". 453–94. Interscience, New York.

Erdtman, H. (1963). Some aspects of chemotaxonomy. *Pure appl. Chem.* **6**, 679–708.

Erdtman, H., Frank, H. and Lindstedt, G. (1951). Constituents of Pine Heartwood. XXVII. *Svensk PappTidn.* **54**, 275–9.

Erdtman, H., Kimland, B. and Norin, T. (1966a). Pine phenolics and pine classification. *Bot. Mag., Tokyo* **79**, 499–505.

Erdtman, H., Kimland, B. and Norin, T. (1966b). Wood constituents of *Ducampopinus krempfii* (Lecomte) Chevalier (*Pinus krempfii* Lecomte). *Phytochemistry* **5**, 927–31.

Erdtman, H. and Norin, T. (1966). The chemistry of the order cupressales. *In* Zechmeister L. (ed.), *Fortschr. Chem. org. NatStoffe* **24**, 206–87.

Erdtman, H. and Wachtmeister, C. A. (1957). Phenoldehydrogenation as a biosynthetic reaction. "Festschrift A. Stoll." 144–65. Birkhäuser, Basel.

Hegnauer, R. (1967). Chemical characters in plant taxonomy: some possibilities and limitations. *Pure appl. Chem.* **14** (1), 173–87.

Leistner, E. and Zenk, M. H. (1967). Incorporation of shikimic acid into 1,2-dihydroxyanthraquinone (alizarin) by *Rubia tinctorum* L. *Tetrahedron Lett.* 475–6.

Liebman, A., Mundy, B. and Rapoport, H. (1967). The biosynthesis of nicotine in *Nicotiana glutinosa* from carbon-14-dioxide. Labelling pattern in the pyrrolidine rings. *J. Am. Chem. Soc.* **89**, 664–72.

Linstedt, G. (1951). Constituents of pine heartwood. A general discussion. *Acta chem. scand.* **5**, 129–38.

Lynen, F. (1967). Biosynthetic pathways from acetate to natural products. *Pure appl. Chem.* **14**, No. 1, 137–67.

Mirov, N. T. (1961). Composition of gum turpentines of pines. *U.S. Dept. Agric. Forest Service. Techn. Bull.* No. 1239.

Mirov, N. T. (1967). "The Genus Pinus." The Ronald Press, New York.

Shaw, G. R. (1914). "The Genus Pinus." Arnold Arboretum No. 5, Cambridge, Mass.

Smith, R. H. (1964). "The Monoterpenes of Lodgepool Pine." *Phytochemistry* **3**, 259–62.

Stebbins, G. L. (1951). "Variation and Evolution in Plants." Columbia University Press, New York.

Westfelt, L. (1967). Terpenes from *Pinus silvestris*. *Svensk kem. Tidskr.* **79**, 3–23.

21. An Assessment of the Value of Biochemical and Serological Techniques in Microbial Taxonomy

S. T. COWAN

Formerly of the Public Health Laboratory Service, Central Public Health Laboratory, London, England

ABSTRACT

From the earliest days, bacteriology has relied on the results of simple biochemical and serological tests by which to characterize different taxa. Newer and more complex methods yield more information, and the analysis of cell walls and the GC (guanine–cytosine) ratios of nucleic acids have done much to confirm and extend classifications made on other grounds. However, the information given by biochemistry and serology is not used in a uniform way to fix the rank of micro-organisms in a hierarchical system, and the relations between characters and categories are compared in several so-called genera (*Streptococcus*, *Salmonella*, *Arizona*, *Escherichia*). Although the experts differ in what they regard as species among the enterobacteria, the majority characterize both genera and species on biochemically-determined characters, while the serological subdivisions are regarded as subspecific serotypes. This comparison leads to a set of guiding principles to serve as a basis for discussion on the correlation of characters with taxonomic rank. Finally, the whole basis of a hierarchical system of classification for bacteria is questioned, and the idea of a spectrum of intergrading forms is welcomed as wholesome food for future thought.

In medicine there is an adage "Use a new drug while it works", and we cannot be blamed if we apply the same scepticism to some of the brilliant results obtained by the sophisticated techniques of modern biochemistry and serology. Before this symposium I made lists of biochemical and serological tests that have been applied during the last seventy or eighty years to microbes to assist in their classification and identification; I tried to keep the two lists separate but, as Professor Westphal's contribution has shown, serology is really a physical

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 269–278.

demonstration of chemical similarity or difference. Briefly, biochemistry has been used to characterize bacteria, to study their nutritional requirements and abilities, their metabolism, and their respiratory preferences, and to identify the chemical nature of structures that make up the cell. Serology also has been a tool for the characterization and identification of bacteria by analysing reactions that take place not only on the cell surface, but with antigenic constituents liberated by disruption of the cell membrane. The metabolic products of some bacteria are pharmacologically active, toxic, and antigenic, and toxin–antitoxin reactions can be used to classify these bacteria. The lists are not exhaustive but give some indication of the uses to which biochemistry and serology have been put in microbial, and particularly bacterial, systematics.

Both serology and biochemistry have helped us to understand the physical and chemical nature of the various morphological structures that make up the bacterial cell. We visualize it as made up of a nucleus containing DNA and perhaps RNA; the cytoplasm may contain globules of fat or fat-like substances, and also enzymes that can be released by mechanical disruption of the cell. Either within the cytoplasm or in the cell membrane are constituents not released by mechanical means but which can be liberated by enzymic action or hydrolysis. Next is the cell wall, made up of chemical elements that can be isolated and identified by chromatography; the physico-chemical nature of this wall determines its reactivity to differential stains such as Gram's method. Sometimes the cell wall contains a capsular polysaccharide that is continually being released or secreted into the surrounding medium. Finally there are the superficial structures such as a slimy covering, finger-like processes called fimbriae, and longer processes (flagella) which are involved in the movement or motility of most bacteria. Cell walls and capsules form what are known as surface or somatic antigens, in serological shorthand they are the O antigens of organisms in the S (originally smooth) form, and their determinant groups can be shown to be on the surface, or exposed, because the intact cell will react with antibody in simple agglutination tests. Cell walls are made up of a basal mucopeptide unit upon which a pattern of mucopeptides and mucopolysaccharides is superimposed, and we know from the work of Cummins and Harris (1956) with the Gram-positive bacteria, and of Westphal with the enteric organisms, that the superimposed pattern is characteristic for a taxonomic group.

Flagella act as antigens, called H antigens in serology and they are independent of S $\rightarrow$ R form variation, and because they are superficial in position, they also are shown by agglutination tests. Deeper antigens such as the group polysaccharides of streptococci cannot be shown by agglutination of intact cells; the cells must be broken up in some way so

that the deeper antigens are released, freed from cell debris to leave the antigen(s) in a clear solution in which it can be detected by precipitation, complement fixation, or cell diffusion methods. The location of antigens has been improved by immunofluorescence techniques, in which antibody is coupled to a fluorescent dye; the site of the antigen–antibody reaction is seen by fluorescence under ultra-violet light. Fluorescence can also be used to locate DNA in microbial cells and to detect parasitization by bacteriophage.

CHARACTERIZATION OF MICRO-ORGANISMS

Only the crudest descriptions of bacteria consist entirely of morphological features and staining reactions, but fungi have a much greater variety of cellular structure and arrangement. Since the earliest days of the discipline, bacteriology has relied on simple biochemical tests to provide the diagnostic and distinguishing criteria. At first, differentiation of species might depend merely on the ability to produce acid in milk or to liquefy gelatin. Many of our early descriptions were based on the results of tests made before we knew how to isolate organisms in pure culture, with the result that the descriptions were not the characters of one organism but the combined (and sometimes interacting or symbiotic) characters of several organisms. In spite of this well-known fact, bacteriologists have been foolish enough to follow the botanists by accepting 1753 as the starting date for nomenclature and they receive almost uncritically any description made after that date. Many of us have protested and we would prefer a starting date in the twentieth century; to effect this would necessitate a change in the Bacteriological Code, and this is the kind of amendment that would raise opposition because, without wholesale conservation of specific epithets, it might involve a large number of changes of name. And although bacteriologists are much more ready to conserve names, both specific and generic, than the botanists, discussions on a change in starting date always get bogged down in detail.

Biochemical characters are used in making the major subdivisions of micro-organisms, such as the divisions into aerobes and anaerobes, into autotrophs and heterotrophs, and classifications have been based on the metabolic products (e.g. acetic acid, lactic acid) or the chemical activities (such as the reduction of sulphur, nitrification). Some bacteria produce toxins that are specific and therefore taxonomically useful, others produce pigments that are pathognomonic, or seem to be so. Antibiotics are produced by some micro-organisms but this may be a strain peculiarity rather than a characteristic of a taxonomic group.

Studies of the nutrition and respiration of bacteria can help in their classification, and one taxon is distinguished at generic level by

dependence on accessory food substances or vitamins which, in nature, may be provided and made available when organisms grow in symbiosis. In this connexion we bear in mind the essential parasitism of viruses and of some bacteria never found outside the cell or the body of the host and apparently incapable of obtaining all their requirements from lifeless media. Tissue cultures are proving useful in growing such organisms in vitro and this should lead to their better characterization.

NUCLEAR MATERIAL AND GENETICS OF MICRO-ORGANISMS

The most characteristic feature of any organism will be imprinted in the nuclear material of the cell, for it is in the nucleus that we have the genetic apparatus, which must determine the maintenance of the organism's specificity. In bacteria and most viruses the genetic material consists of DNA, and in some viruses of RNA. The GC ratio (guanine + cytosine: all bases) is significant not only to the geneticist (for it is those bacteria with the greatest similarity of GC ratio that are most likely to hybridize) but also to the taxonomist, for the production of hybrids is surely the best of all tests of relatedness. However, similarity of GC ratio does not necessarily imply relatedness but, more significantly, dissimilar ratios show complete absence of genetic similarity and, therefore, indicate unrelatedness.

Experimental hybridization and other evidence of the passage of genetic material from one organism to another may be found in bacteria that are obviously very similar and probably closely related. Since some of the hybrids are crosses between so-called genera, this gives us reason to question the basis of our generic differentiation. Some of the artificial hybrids are very like natural groups that are awkward to classify and seem to occupy positions intermediate between more clearly defined taxa; members of a naturally-occurring group of bacteria known as the A–D group (meaning alkalescens-dispar, epithets of species that were once accepted) are intermediate between *Escherichia* and *Shigella*, and these organisms very closely resemble the artificially-produced hybrids of *Escherichia* and *Shigella*. In the same way, the naturally-occurring *Arizona* cannot be distinguished from hybrids of *Salmonella* and *Escherichia*. Although *Escherichia*, *Shigella*, *Salmonella* and *Arizona* are usually regarded as genera, the experts who work with these organisms are by no means unanimous about their classification, and the debate is long, detailed and acrimonious.

SPECIFICITY OF SEROLOGICAL REACTIONS

The short answer to the question of the specificity of serological reactions is that they are not specific; an antiserum will react not only

with the homologous antigen(s) but also with partial antigens (haptens) that have some of the same or very similar determinant groups; for example, a pneumococcus type II antiserum cross-reacts with the capsular antigen of Friedländer's bacillus type B (which is a very different kind of bacterium from the pneumococcus) and also with a polysaccharide from gum arabic.

When cross-reactions occur between heterologous antigens and antibodies we have to distinguish between the sharing of the same antigen by two different organisms and cross-reactions due to haptens and non-specific factors. Absorption will usually make the position clear but it may be necessary to do mirror tests and to test the antigenicity of each organism. An apparently single reaction in some of the cruder methods (e.g. ring precipitin tests) may, by the optimal proportions method or by agar diffusion, be shown to consist of several antigen–antibody systems reacting in the same mixture, and it is obvious that the more elaborate techniques will yield the most information. We have known for a long time that organisms may share antigens, and when they share the same antigens in different proportions it is a fair assumption that the organisms are related. The complex sharing of antigens among the salmonellas and other members of the Enterobacteriaceae is so widespread that it is difficult, on serology alone, to say more than that these bacteria are (untaxonomically speaking) a "happy family". But we have other and simpler examples: the bacteria that cause brucellosis share two antigens, labelled A and M, in different proportions; the one that causes undulant fever in the Mediterranean area has much M and little A; the one that causes undulant fever in man in this country and abortion in cattle has mostly A and a little M; the pig organism has only slightly more A than M. These bacteria have many cultural characters in common and there is continual discussion about their rank; are they three separate species or three varieties of one species, *Brucella melitensis*? There are die-hards on both sides and serology alone cannot settle the argument.

NUTRITION, METABOLISM AND EVOLUTION

Stanier *et al.* (1966) described studies on pseudomonads in which both the nutrition and the metabolism of the organisms played a big part in their classification. When an organism needs a particular substance for its nutrition or as an accessory food factor, it is reasonable to argue that it cannot synthesize that substance; from this it is easy to speculate on the relative positions of the microbial gourmands and gourmets on an evolutionary scale. Taxonomists who believe that a classification should indicate evolutionary trends will be attracted to schemes based on

nutritional requirements but, as nearly always, there are at least two ways of interpreting known facts; for example, Knight believes that evolution followed a line of increasing complexity of nutritional demands, while van Niel reverses the order, and regards the photosynthetic organisms as the most highly developed with the least dependence on external reducing factors.

Within recent years bacteriologists have paid more attention to what might be called the respiration of the organism, particularly the way in which it assimilates glucose, and a convenient subdivision of bacteria can be made into "oxidizers" and "fermenters". This leads to a practical difficulty; microbiology being an applied science, most microbiologists had a basic training in some other subject such as botany, zoology, medicine or engineering, and few of the older members have a good enough grounding in chemistry to take advantage of all that the biochemist can offer. It is essential, therefore, that relatively simple and preferably rapid methods should be developed; there is a simple test for the oxidation/fermentation character (Hugh and Leifson, 1953) but it can be applied only to organisms that can grow in the basal medium and other methods must be used for more fastidious bacteria such as the *Neisseria* spp.

SEROLOGY AND BIOCHEMISTRY IN FIXING RANK

In general, the high ranks (such as order) of bacteria are distinguished by morphology (i.e. cytology and staining reactions) and by major physiological features such as photosynthetic pigments. At lower levels (family) differentiation is based on the type of respiration, metabolism and metabolic products. Within a family, generic differentiation is occasionally based on morphology (as in *Nitrobacteraceae*) but more often on additional physiological and metabolic differences. Speciation is made on the minutae of biochemical differences; sometimes habitat and pathogenicity are introduced, but in microbiology these are not very useful taxonomic characteristics. Subspecies may be defined on even smaller biochemical differences (e.g. the fermentation of a single sugar) or by serology and/or phage sensitivity when these are applicable, as, for example, in *Staphylococcus*.

Bacteriologists are blessed (if that is the right word) with only the vaguest ideas on taxonomic rank, and the species in one genus is not necessarily comparable to species in another. I will illustrate this by describing and comparing the subdivisions of some so-called genera in two families (*Lactobacillaceae* and *Enterobacteriaceae*) that have been well characterized both by serology and by biochemistry. To keep the picture relatively simple I will confine the comparison to *Streptococcus*

in one family, and to *Salmonella, Arizona* and *Escherichia* in the other.

Streptococcus is divided into species partly on biochemical (B) differences and partly on serological (S) groups based on the Lancefield precipitin tests made with polysaccharides from the hydrolyzed bacterial surface. These serological groups are labelled by letters, but new groups are continually being found and the alphabet will soon be exhausted. Most of these serological groups have different major biochemical characteristics, and some of the groups correspond with species defined earlier in biochemical terms. In other serological groups (of the Lancefield series) there are biochemical subgroups, and these are sometimes graded as species, and sometimes as subspecies or varieties. Comparable to the Lancefield groups, and distinguished from them by another polysaccharide antigen, is a species, *Streptococcus pneumoniae*, commonly called the pneumococcus. Finally, serologically and biochemically distinct, but not yet sorted out, is another and heterogeneous group which is given the honorary title of *Streptococcus viridans*.

Of the groups labelled by letter, A has been most thoroughly studied, followed closely by B, C and G (all these can be human pathogens). Group A has the species name *S. pyogenes*, and the taxon is subdivided into numerous serological types by the M and T antigens. Group B (*S. agalactiae*) is subdivided by precipitin and agglutination tests into four main types and several subtypes. On biochemical characters group C is divided into four species (*S. equi, S. equisimilis, S. zooepidemicus, S. dysgalactiae*) and is further subdivided into serological subtypes. The enterococci and *S. bovis* form group D; there are about 40 serological types of the enterococci, which are also divided on biochemical grounds into several species and/or varieties.

I do not intend to go down the alphabet; enough has been said to show that the treatment of the different taxonomic groups is diverse. *S. pneumoniae* is not in the lettered series; indeed, in the U.S.A., largely because of the great influence of "Bergey's Manual of Determinative Bacteriology" (1923–57), this is held to belong to a separate genus (*Diplococcus*). The pneumococcus has a polysaccharide capsule which determines its serological type, and there are about 50 types.

In the family Enterobacteriaceae the so-called genera are distinguished by multiple but minor chemical reactions or characteristics; the O antigens are shared both among genera and divisions of genera. There are several hundred known (and naturally occurring) combinations of *Salmonella* O antigens and the permutations seem endless; at one time, when the discovered combinations were fewer in number, each combination was regarded as a species; Kauffmann still holds this view but the majority of those who now work with the *Salmonella* group treat the different antigenic combinations as serotypes or subspecies. *Salmonella*

O antigens are shared with *Arizona* (which itself is variously regarded as a subgenus or a species of *Salmonella*), and with *Escherichia*. There are about sixty serotypes of *Arizona*, and more than a hundred of *Escherichia*, in which only one species is now recognized.

The parts played by biochemistry (B) and serology (S) in the characterization of the different ranks are shown in Table 1, and the subdivisions made possible by bacteriophage (*P*) typing have been added. It is clear that here is a fertile field for discussion of the meaning of taxonomic rank in bacteriology. The first question that arises is: should the species and subspecies of one genus be comparable to the species and subspecies of another genus? If we think that they should be comparable, are the hundreds of serotypes of *Escherichia*, of *Salmonella*, and of *Arizona* to be given the status of species, or should they, like the serotypes of the pneumococcus or of group A streptococci, be regarded as subspecies? Dr. Westphal has told us that the third component of the deeper antigens of different "genera" among the Enterobacteriaceae may differ, and this new evidence is bound to influence our thinking.

TABLE 1. The main characteristics of taxonomic rank in some bacteria

Rank	*Streptococcus*	*Salmonella*		*Arizona*	*Escherichia*
		Kauf	Others		
Genus	B	B	B	B	B
Species	B and/or S	SBP	B	B	B
Subspecies	S/B	.	S	S	S
Sub-subspecies	S	.	P	.	.

Kauf = according to Kauffmann
B = *biochemical characters*
P = bacteriophage typing
S = serological characters or antigenic structure

If we assume that taxa of the same rank are to be comparable we should be able to draw up lists of the qualities that distinguish the different ranks, but this is not an easy task and may be an impossible one. In its place I have drawn up a set of guiding principles that could form the basis for discussion.

1. Groups with the same basic biochemical characterization should form a genus.

Such a genus might be characterized by its nutritional requirements, respiration (aerobic or anaerobic), metabolism (assimilation of glucose), and taxonomically useful characters such as catalase and oxidase.

2. Genera should be distinguished by the possession of one or more uncommon (preferably unique) characters.

The distinguishing features will consist of the mucopeptides and mucopolysaccharides of the cell wall. At this stage can be included the DNA base ratio.

3. Within the genus, species should be distinguished by multiple minor differences in biochemical characterization.

The kind of biochemical differences I have in mind here is that shown by simple everyday tests used in all laboratories in the identification of bacteria, tests that do not need to rely on the expert help of the biochemist; examples are the tests for indole, nitrate reduction, liquefaction of gelatin. When these differences are supported by differences in serology (e.g. the Lancefield grouping of streptococci), so much the better, as the addition evidence lends weight to the speciation so made.

4. In general, serological subdivisions should have subspecies status. This needs qualification because some antigens, especially those deep in the cell and not detectable by agglutination tests, may be shared on a broader, possibly generic, base. We know that the pneumococcus has deep protein and polysaccharide antigens that are found also in other streptococci. Taxonomic groups that share the same surface antigens (H and O) in different proportions or different combinations should have the status of subspecies.

Most bacteriologists are splitters and it is extraordinarily difficult to find the bases on which the different taxonomic ranks rest. A little rationalization would produce a good deal of clumping and ultimately, a happier breed of bacteriologist. But before that utopian state arrives we shall have to consider seriously the suggestions that bacteria cannot be put into hierarchical schemes of classification and that, anyway, not more than two ranks can be justified. Kauffmann (1954) described the Enterobacteriaceae as being "made up of interrelated bacteria which do not lend themselves to sharp divisions", and Waksman (1957) wrote of actinomycetes that "the species are so gradually linked that it is very difficult to say where one species ends and another begins". Abd-el-Malek and Gibson (1948) thought that staphylococci and micrococci could be arranged in a virtually continuous series, a concept modified by Shaw *et al.* (1951) by likening it to a spectrum.

This idea of a spectrum expresses both the biochemical and serological characteristics of bacteria. The DNA base composition charts published by Hill (1966) go even further and suggest that when bacteria are arranged (by % GC) in a spectrum or series, the clusterings and separations make logical sequences. While it is obvious that all organisms with the same GC ratio cannot be related, it is possible that a series of

GC charts could be made to bring out spectra to reveal new taxonomic groupings.

To arrange bacteria in a spectrum or a series of spectra to include bacteria of all kinds, would mean the abolition of a hierarchical scheme and a complete break with classical taxonomic practice. For a long time bacteriologists have grown restless with the restraint imposed by classical nomenclature, based as it is on a hierarchical system of classification, and there is at present a great rethinking of the whole problem by those whose work brings them closest to the problem. All are aware that serology and biochemistry have contributed much to microbial taxonomy and, together with microbial genetics, must form the basis of any classification, whether it takes the form of a hierarchical pyramid or a linear spectrum.

REFERENCES

Abd-el-Malek, Y. and Gibson, T. (1948). Studies in the bacteriology of milk. II. The staphylococci and micrococci of milk. *J. Dairy Res.* **15**, 249–60.

Bergey's Manual of Determinative Bacteriology (1923–57). Seven eds, 1923, 1926, 1930, 1934, 1939, 1948, 1957. Williams & Wilkins, Baltimore.

Cummins, C. S. and Harris, H. (1956). The chemical composition of the cell wall in some Gram-positive bacteria and its possible value as a taxonomic character. *J. gen. Microbiol.* **14**, 583–600.

Hill, L. R. (1966). An index to deoxyribonucleic acid base composition of bacterial species. *J. gen. Microbiol.* **44**, 419–37.

Hugh, R. and Leifson, E. (1953). The taxonomic significance of fermentative versus oxidative metabolism of carbohydrates by various Gram negative bacteria. *J. Bact.* **66**, 24–6.

Kauffmann, F. (1954). "Enterobacteriaceae." Ejnar Munksgaard, Copenhagen.

Shaw, C., Stitt, J. M. and Cowan, S. T. (1951). Staphylococci and their classification. *J. gen. Microbiol.* **5**, 1010–23.

Stanier, R. Y., Palleroni, N. J. and Doudoroff, M. (1966). The aerobic pseudomonads: a taxonomic study. *J. gen. Microbiol.* **43**, 159–271.

Waksman, S. A. (1957). Species concept among the actinomycetes with special reference to the genus *Streptomyces*. *Bact. Rev.* **21**, 1–29.

22. Chairman's Summing-Up

J. Heslop-Harrison

Institute of Plant Development, University of Wisconsin, Madison, Wisconsin, U.S.A.

Traditionally a chairman is allowed to begin his summarizing remarks with a comment on the difficulty of his task, and after a symposium as wide-ranging as this one, I suppose I would be justified in declaring the job practically impossible. Our organizers have brought together an outstanding cast of contributors from many different fields, and these contributors have expounded techniques and discussed results of a diversity beyond the capacity of any one to assimilate and evaluate on a first hearing. However, the common thread is the relevance for taxonomy, and it is to this aspect I propose to address my remarks. I am sure that in the minds of many taxonomists in the audience the central question concerns the true significance of the new sources of comparative data for their trade. Can it be said that any of the novel criteria have in consequence of their nature such special value for classificatory purposes that they transcend in significance those that have hitherto provided the main basis for the classification of organisms? Or should taxonomists accept the new data sources as welcome addenda to those already available, and treat the evidence they offer in essentially the same manner?

In commenting on this problem at the first International Symposium on chemical plant taxonomy in Paris five years ago, I emphasized that there was in essence no quality inherent in chemical data differentiating them as a class from others available as classificatory criteria. Both chemical and morphological properties of the phenotype are the outcome of biosynthetic processes governed by the genome, and so far as it can be said to be the taxonomist's aim to classify genotypes (a point to which we will revert), he is likely to find as reliable guides in the one as in the other. At the time of the 1962 meeting we were discussing primarily the classificatory value of the data accruing from plant

Systematics Association Special Volume No. 2. "Chemotaxonomy and Serotaxonomy", edited by J. G. Hawkes, 1968, pp. 279–284.

product chemistry. In this meeting we have been delving still further, and considering not only the final products of biosynthetic pathways but more direct manifestations of the genome, and indeed the chemistry of the genome itself. Evidently the question needs reconsideration, lest it indeed be true that some of the newer approaches have such overwhelming power and precision that they are destined to revolutionize taxonomy.

The scheme set out in Fig. 1 shows some of the relationships between the classes of compounds considered in this symposium, and places them

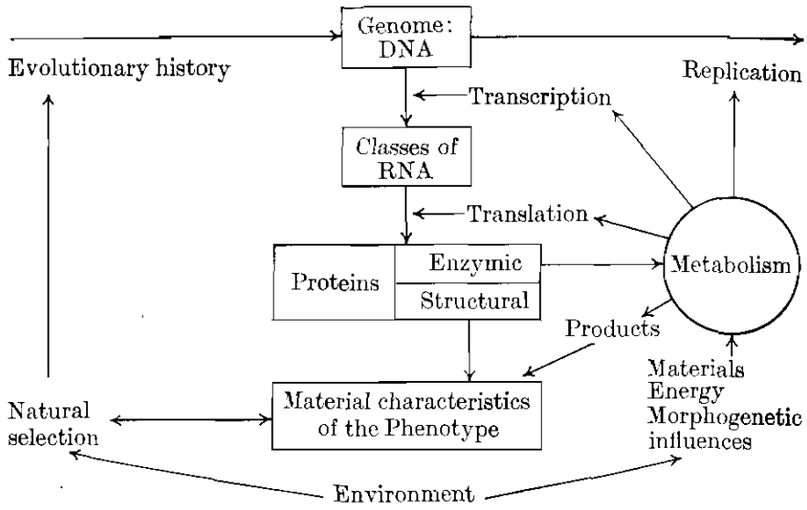


Fig. 1

FIG. 1. Inter-relationships of classes of compounds with potential taxonomic usefulness. DNA, some classes of RNA and proteins are information-carrying polymers. The micromolecules useful as taxonomic discriminants are products of metabolism; their presence in cells and tissues must indicate the existence of biosynthetic pathways governed by specific enzymes, and these in turn must be coded for in the genome. The environment is shown as affecting (1) the genome, through the feed-back loop provided by natural selection, and (2) the ontogeny of the individual, through the supply of materials and energy and the variety of secular influences directly influencing development.

in the biochemical hierarchy. A distinction one may immediately make in considering this scheme is between macro- and micromolecules, or perhaps more precisely between polymeric and non-polymeric molecules. Dr. Harborne has suggested that this distinction may not be of great importance in the present context. In respect to biological role, however, it is the difference in potential for information storage which

is profoundly significant. Nucleic acids, proteins, perhaps some polysaccharides, carry information in the form of a quasi-alphabetical code, and their flexibility and capacity in this role is potentially unlimited. In contrast, diversity in the non-polymeric molecules is limited; the number of "bits" of information that can be carried in the structure of the molecule is really quite small, even with such classes as the steroid and gibberelloid hormones, and in any event it seems always necessary for there to be a polymer in the cell to recognize and translate such information as these small molecules might convey. In studying macro- and micromolecules in relation to taxonomy we are therefore concerned with quite different things. The small molecules are assuredly present or absent because their synthesis is either coded for or not in the genome, but they do not reflect the genetic code directly; in contrast, the proteins represent translations of the code itself.

However, it certainly is true that the distinction has not been very meaningful in the use made of chemical information hitherto in taxonomy. In practice, this has involved the seeking out of resemblances and differences, often on the simple basis of "presence" or "absence". For the smaller molecules this has depended upon identification and usually, although not always, full characterization; and for proteins, comparison of certain properties such as molecular size and electrophoretic mobility, immunogenic properties and, in a few instances, fingerprints and amino acid composition. In all cases the particular molecules to be compared have been selected because they do show variation within the major group under investigation, and as Dr. Bate-Smith has pointed out, this has meant the selection of metabolically inert molecules of considerable structural diversity. Without entering too deeply into discussion of what are and what are not adaptive characters, one may say that in general this must mean molecules that are either in some sense dispensable for the organism, or if not dispensable, then free to vary within certain limits structurally without prejudicing biological function. The chemical properties adopted for taxonomic purposes thus represent *fortuitous* characteristics, in the sense in which Wernham applied the description to morphological features useful in taxonomy because they were not too strongly correlated with the capacity for survival itself.

Perhaps more clearly than ever before we are shown the meaning of "fortuitous" variation in the facts concerning haemoglobin structure discussed by Dr. Huntsman. The shape of the molecule is the determinant of its function as a respiratory pigment, and this is determined in turn by the amino acid sequence. But at nine sites a substitution is tolerable because the pattern of folding is not affected. Diversity at these nine "free" sites then provides the inter- and intra-group

variation that assumes taxonomic—and phylogenetic—significance. Presumably the same principle applies in large measure to the serum and storage proteins discussed by Professors Gell and Hawkes, to Professor Blombäck's fibrinogens, and to the ribosomal RNAs of Dr. Gould. The discriminating characteristics these investigators have been able to detect are probably for the most part in properties of the molecule where variation is permissible without prejudice to function. Where this is true, the accumulated variation may preserve the record of accidental evolutionary change, and provide some of the evidence for reconstructing part at least of the réseau of evolutionary relationships. We may note one special advantage of studying such fortuitous features in macromolecules, namely that they are likely to possess that great merit in the eyes of taxonomists, consistency in expression. If a particular protein, let us say, has a biological role, it will be present in all individuals of a lineage; and if an "error" is coded in it by the genome, this error will also be expressed in all individuals. This is not necessarily true with micromolecules of the metabolically inert type mentioned by Dr. Bate-Smith. The enzymes concerned in the biosynthetic pathways are genome determined, but whether the pathways function is something established during development. So far as non-essential products are concerned, substantial environmentally-conditioned variation will almost always occur, since the less vital a function is to survival, the less likely that it will have become protected against chance fluctuation by the establishment, under selection, of buffering gene systems.

Although we may suppose that within comparatively narrow circles of affinity it is chemical variation of the "fortuitous" type that is likely to provide evidence of the greatest taxonomic value, this is not to suggest that chemical differences of obvious functional significance may not offer themselves as criteria. Professor Boulter has mentioned isoenzymes as providing one dimension of variation in plant protein extracts. Changes in isoenzyme patterns commonly accompany differentiation, and they may thus be tissue specific; correspondingly, it is obviously a possibility that related taxa with similar overall metabolism could reveal enzymic heterogeneity in this same manner. The case is different only in degree from that of the haemoglobins: variation in the enzyme complements, and in the DNA base sequences determining them, is tolerable while the substituents continue to discharge their biological function.

When comparisons are made between more remote groups of organisms, it is to be expected that, as with morphological features, chemical differences will be apparent at more profound levels of organization—and not only will the differences be more profound, but there will be more of them. This is the natural outcome of the lesser

community to be expected between the genomes of groups with long independent evolutionary histories. It is noteworthy that such differences may be attributable not only to divergent adaptive evolution, but to the erosion of similarities in the heritage from the last common ancestor by fortuitous changes in genome code not directly concerned with adaptive change. Here we have the possibility—indeed probability—of a curiously paradoxical situation, where the phenotypes provide better guides to evolutionary affinity than the genetic codes themselves. Mitochondria look much the same, and function in essentially the same way, in widely diverse organisms: but who is to say whether their structural and enzyme components are *now* encoded in the same base sequences in the nuclear or mitochondrial genomes of all groups with these organelles? Indeed, what evidence there is suggests that this is not so. Selection—both progressive and stabilizing—is for the phenotype, and therefore only indirectly for genome coding; if the same phenotype can be coded by different sequences, then these will all be equally acceptable. It is exactly as with polygenically determined quantitative variation: if the same phenotypic expression can be established by different polygene combinations, selective factors in the environment cannot distinguish them, and all will be equally viable.

At this level of analysis, the definition of "homology" given by Dr. Gaylord Simpson as "resemblance due to inheritance from a common ancestry" loses much of its meaning, and the distinction between homology and analogy fades away. Perhaps more disturbing is the implication that the commonly accepted aim of classifying "genotypes" is not necessarily the most appropriate one for taxonomists, even could it be achieved, when a genomic *difference* may have less meaning as an index of "degree of affinity" than a phenotypic *resemblance*.

These considerations lead me to feel that my conclusion of 1962 really needs no amendment; whatever their source, chemical data, as taxonomic criteria, are to be treated as indicators of similarity or difference in the particular context of each group, and evaluated on the same basis as any other characteristics of the organism. It seems to me that this conclusion applies as well to the evidence of DNA hybridization discussed by Dr. Kohne as to any other; DNA homology provides simply one more special test of affinity, comparable, say, to genetic hybridization, and does not offer a definite assessment of "similarity" of transcendental taxonomic importance. Since resemblance at this level is not obligatorily correlated with resemblance at all other levels, this would be so even if the genome were laid out as linear, non-repetitive sequences, and if the experimental techniques were such as to give unambiguous measures of identity—conditions certainly not fulfilled for higher organisms.

This conclusion is, of course, far from being a rejection of chemical data as being irrelevant for taxonomy: if anything is certain it is that they will in the future provide an increasing part of the evidence available to the working taxonomist in all groups. As such they have many potential advantages, including for some classes of compounds the particular merits of consistency, ease of assay and unambiguity; and this aspect has emerged particularly clearly in the present symposium from Dr. Erdtman's review of the recent history of chemical plant taxonomy. We have also the lesson of the microbial groups before us. The contributions of Dr. Norris and Professors Westphal and Stanier, and the review of Dr. Cowan, show that chemical data now form more than ever the principal source of criteria for micro-organism classification—as indeed is to be expected, since they constitute the most readily observable expressions of the genome with organisms revealing so little in the way of structure. It is improbable in the extreme that comparative morphology should ever be superseded as the principal source of taxonomic criteria for the higher organisms, but if the proper aim of taxonomy is to produce classifications based upon the maximum correlation of attributes—Professor Cain's phenetic classifications—it is inevitable that chemical data will contribute increasingly. Whether this will lead to a massive re-shaping of existing classifications is an open question. My own opinion is that should this happen, it will only be because the chemical data have been misused. The most satisfactory parts of our present classifications are based upon the correlation patterns shown by many attributes of organisms. The availability of new criteria is bound to alter the emphasis in some cases, particularly where existing taxonomy has been founded upon inadequate evidence; but for the better-known groups it is to me inconceivable that the addition of another random sample of genome-expressions should alter in any radical manner the patterns of similarity and difference manifest in the sample already taken into consideration in traditional taxonomic practice.

Author Index

Numbers in italic indicate the page on which the reference is listed.

T

U

V

W

Subject Index

Comparative Phytochemistry

Edited by T. SWAIN

Low Temperature Research Station, University of Cambridge, England

1966, xiii + 360 pp., 93s.

Comparative phytochemistry deals with the distribution and biogenesis of chemical substances in plants. The subject is now well established and has proved of interest not only to plant taxonomists, but also, by indicating interesting taxa and biosynthetic sequences, to natural product chemists.

This volume, based on papers presented at a NATO sponsored meeting of the Phytochemical Group in Cambridge, can be regarded as an important up-to-date account of the present state of the study and the progress made in this field in the last few years.

"*The book deserves a wide readership; much otherwise scattered information is gathered in one volume and the book will indicate to workers in several fields the progress being made in areas adjacent to their own.*"

Chemistry in Britain

Contents

Academic Press
London
and New York

Berkeley Square House,
Berkeley Square, London, W1X 6BA

111 Fifth Avenue, New York,
New York 10003

Printed in England