Biochemical Evolution

EDITED BY

H. GUTFREUND

Biochemical Evolution

EDITED BY

H. GUTFREUND

PROFESSOR OF PHYSICAL BIOCHEMISTRY
UNIVERSITY OF BRISTOL

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE

LONDON NEW YORK NEW ROCHELLE

MELBOURNE SYDNEY

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1981

Printed in Great Britain by
J. W. Arrowsmith Ltd., Bristol BS3 2NT

British Library Cataloguing in Publication Data

Biochemical evolution.
1. Chemical evolution
I. Gutfreund, Herbert
575 QH325 80-49949

ISBN 0 521 23549 9 hard covers
ISBN 0 521 28025 7 paperback

Contents

1

Some problems in molecular evolution

H. GUTFREUND
DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF BRISTOL, BRISTOL BS8 1TD, UK

Nothing in biology makes sense except in the light of evolution (Dobzhansky, 1973).

In recent years there has been a considerable increase in the literature on many aspects of the study of biological evolution. This has ranged from philosophical (Popper, 1972) to mathematical, functional and chemical aspects (J. M. Smith *et al.*, 1979). Even restriction of the discussion to biochemical evolution requires considerable selection of topics from a great range of possible approaches to present some of the salient ideas in a small volume. Since the present work is intended to interest and instruct biological scientists including honours and graduate students, it is important to emphasise that for each conclusion drawn here, from the facts presented, there are likely to be very different views expressed by other authors. Even among the authors of the present volume alternative or conflicting proposals were put forward, sometimes by the same author and retained by the editor. Any attempt to discuss more than a small selection of the possible interpretations of the evidence would confuse rather than educate and stimulate the reader.

Biochemical evolution can be treated in three inter-related ways. Firstly, one can think about the evolution of chemical reactions occurring in biological systems from the prebiotic processes discussed by Schuster in chapter 2, to the catalysis and control of enzyme mediated metabolic sequences discussed by Clarke in chapter 4. Secondly, one can use biochemistry as a tool to study the sequence of events in evolution. Chapter 3 on computer methods for the construction of phylogenetic trees is dependent on the enormous amount of information which is becoming available from ever-improving biochemical techniques for the analysis of amino-acid and nucleotide sequences. To varying extents biochemical techniques and approaches have provided key

information about the evolution of different physiological functions, of which only a small and necessarily arbitrary selection can be included in this volume. In this respect modern biochemistry is an equal partner with palaeontology and other classical biological subjects in its contribution to the sum total of our present knowledge about the sequence of evolutionary events.

A third aspect of biochemical evolution, neglected here because it could easily fill a separate volume, is the use of arguments from the theory of evolution as a tool to elucidate biochemical pathways and mechanisms. Krebs (1979) in a splendid article entitled 'On asking the right kind of question in biological research' applies the rule that non-functional characters do not survive in the course of evolution. He states this as 'a rule and not a law. A law has no exceptions but a rule may have exceptions'. The applicability of this 'rule' is likely to depend on whether one considers evolution at the level of function or near the gene (single bases or amino acids).

The fundamental problems of biochemical evolution 'evolve' into those of morphological evolution. In prebiotic and prokaryotic evolution the competition is first between single molecules in their efficiency to replicate, utilising the available natural products. In bacteria one can still define the evolutionary advantages of single mutations and primary gene products (enzymes) which enable the organism to live in a new environment. We are, however, already confronted with the problem of, often quite long, sequences of metabolic pathways. The possible backwards evolution of these pathways is discussed by Clarke in chapter 4. As the complexities increase during the evolution of higher organisms the competition is more and more between functions determined by many genes. However, even in vertebrates a single mutation can be lethal if, for instance, it results in a non-functional haemoglobin. Alternatively a mutation can enable animals to spread into a new niche if they become fitter in an environment of lower oxygen tension.

Discussions about the theory of evolution have often been confused by the introduction of beliefs instead of reason and facts and by lack of care in the language used to express the ideas. The essentially stochastic nature of evolution is a gamble where the winner takes all. It is usually argued that in a functional complex organism a random change in a gene (mutation) will most likely be deleterious. On rare occasions a mutation will result in off-spring fitter for their environment which will produce more progeny than other individuals. This increased fecundity can be brought about in a variety of ways, such as longer reproductive life,

greater success in attracting mates or greater fertility. It is important to emphasise that the environment has no influence on the probability of a mutation which would adapt the organism to this environment. Of course, radiation and chemicals can cause mutations, but these will cause mutations at random and not for the benefit of the organism. There are mechanisms for adaptation to the environment, but they rely on the genetic history of the organism which has provided it with the necessary control function to switch from the utilisation of one metabolite to another by the turning off and on of genes for the synthesis of different enzymes (see chapter 4). Darwinian evolution functions through mutation, reproduction and selection. A detailed discussion of the principles involved will be found in a number of recent elementary treatments (Smith, 1975) or popular reviews (for example, *Scientific American*, September 1978). The interested reader not familiar with the principles of genetics is also well advised to study an introductory text or relevant chapter in a general biology volume. In modern treatments (for instance, Suzuki & Griffiths, 1976) one finds the molecular as well as the biological basis of mutation, reproduction and inheritance. Another more general survey of Darwinian evolution and critical discussion of its many fallacious interpretations is found in a collection of essays by Gould (1979). It is important to emphasise that Darwin (1809–82) did not know of Mendel (1822–84) and his laws of inheritance.

In the twentieth century Darwinism, relying on the principles stated above, was fused first with population genetics and then with molecular genetics to form what is often termed Neo-Darwinism. Darwin, from the information available in his time, considered a more or less common gene pool throughout a given population. The advent of the 'one gene, one enzyme (one protein)' hypothesis and the sensitive new methods for the detection of single amino-acid displacements (mutations) in proteins have made it possible to add to Darwin's evolution by natural selection. Lewontin (1974) and Harris (1975) described in detail how the electrophoretic analysis of enzymes established the large degree of allelic variation in *Drosophila* and human populations respectively. In this chapter I shall discuss briefly two aspects of this extended study; first the principles and consequences of polymorphism and secondly the evolutionary clock.

With the information now available one has to consider four causes of evolution.

(1) Mutation pressure, which is a slow process resulting in the change of only about 10^{-5} gametes per generation.

(2) Genetic drift, which only happens in small populations, where large chance-fluctuations of gene frequencies may occur.
(3) Gene flow, where immigrants introduce alleles as in mutation pressure.
(4) Natural selection by fecundity through length of reproductive life, survival of the young, mating preference, fertility etc. Many of these factors are related to adaptation to the environment, including food supply.

A large population without mutations or immigration will not evolve. It will rapidly reach an equilibrium distribution of genes according to the Hardy–Weinberg law (see for instance, Suzuki & Griffiths (1976) *p.* 407). The reader should compare these comments with those made in the next chapter (p. 75) on 'four prerequisites for Darwinian evolution'.

The essays presented in this volume are only a small selection of topics which might interest the biochemist or are based on biochemical studies contributing to our knowledge of evolution. It was felt by the editor that a few topics treated in some depth would be more stimulating than many short treatments fitting into a small volume. A very important aspect of evolution, which does not yet receive the attention of the biochemist, but should just be mentioned, is speciation: the multiplication of species. The reader is referred to Mayr (1963) for a detailed discussion of this topic.

Function rather than structure has been emphasised and the reader should be briefly referred to some subjects of interest which have been neglected. In recent years, for instance, there has been a very active debate on the evolution of structural domains which are found as common features in different enzyme systems. The principles of divergent, convergent and parallel evolution are discussed by Peacock in chapter 3. In arguments about the evolution of three-dimensional molecular structures, one has to include the question whether common structural domains in a group of enzymes are primarily of functional significance or whether they are due to a limited number of stable configurations in proteins (Matthews, 1977).

Another interesting subject which is only briefly discussed (chapter 5), concerns the evolution of eukaryotes from prokaryotes and especially the biochemical studies of the origin of mitochondria (see, for instance, Schatz, 1979; D. C. Smith *et al.*, 1979). Some recent findings of profound differences in transcription in the present forms of the two types of cells have been used to argue against the hypothesis of the symbiotic evolution (Darnell, 1978). Another aspect of biochemical studies in this area is the comparison of the subunit structures of enzymes of gram-negative and

gram-positive bacteria with those of mitochondria (Henderson, Perham & Finch, 1979).

The problem of explaining the evolution of a process involving a sequence of steps, each determined by a single gene, has already been referred to and is discussed further in chapter 4 in connection with metabolic pathways. The explanation may be different in the case of the physiological functions with which some of the other chapters are concerned. For the student of evolution 'survival of the fittest' has worked too well. The less efficient mechanisms have not survived for our perusal. The essay by Crescitelli could only cover recent aspects of the perfection of the mechanism of photoreceptor response (chapter 9). All visual systems studied so far depend on the photoisomerisation of retinal for light detection and on small differences in the protein opsin, to which the chromophore is bound, for different spectral sensitivities. Three main branches of the animal kingdom – molluscs, arthropods and vertebrates – have developed eyes which are anatomically profoundly different. It seems that the various anatomical, i.e. optical, arrangements will serve their different special purposes, but a particular principle of photochemistry was universally accepted. The ionic and chemical mechanisms involved in the transmission of the signal from rhodopsin to the optic nerve remain to be elucidated. Much of the specialised evolution may be found there, when these processes are compared between species along the phylogenetic tree.

The problem is a different one when the proteins involved in motility are considered. Actin and myosin seem to be used for many different purposes in different animals. This topic is treated in detail by Weeds & Wagner in chapter 8 and it gives some insight into the evolution of systems which are subject to functional control by calcium ions.

It seems quite possible that some steps in complex sequences may have been useful on their own for another purpose. Enzymes may have been less specific and thus able to catalyse more than one reaction or to take part in cascade mechanisms; for example, in the formation of fibrin clots, one of the many serine proteases involved may have been able to fulfil the function of all of them, albeit less efficiently.

POLYMORPHISM

Studies of the rate of mutation, to be discussed below, have shown that a population consisting of several million individuals has a reasonable chance of some mutations in most genes represented in the population.

Although most of these mutations are maladaptive, when causing major changes, many of them will have minor consequences and may be neutral. Electrophoretic studies mentioned above suggest that normal populations have a large pool of genetic variation.

Clearly a rapid analytical method has to be used for a large population survey of different polymorphic forms of a particular enzyme from many individuals. Complete sequence analysis is used to construct phylogenetic trees and the time scale of evolution (the evolutionary clock discussed in a later section), but it is too time-consuming for the large number of samples to be analysed in a search for the number of mutations in a particular protein in a defined population or species. For such screening of enzyme samples the most widely used technique is electrophoresis. This method will, of course, only detect amino-acid substitutions involving changes in charge. About one third of all amino-acid substitutions would be expected to be of this type. This is only one of a number of factors which would cause one to under-estimate the number of mutations from comparison of proteins by electrophoresis. The main factors are given below.

(1) The degeneracy of the genetic code (Table 1.1) resulting in single base changes without amino-acid substitution.
(2) The production of proteins without activity, resulting in undetected translation or silent alleles.

Table 1.1. *The genetic code*

1st↓ 2nd →	U	C	A	G	↓3rd
U	PHE	SER	TYR	CYS	U
	PHE	SER	TYR	CYS	C
	LEU	SER	STOP	STOP	A
	LEU	SER	STOP	TRP	G
C	LEU	PRO	HIS	ARG	U
	LEU	PRO	HIS	ARG	C
	LEU	PRO	GLUN	ARG	A
	LEU	PRO	GLUN	ARG	G
A	ILEU	THR	ASPN	SEB	U
	ILEU	THR	ASPN	SER	C
	ILEU	THR	LYS	ARG	A
	MET	THR	LYS	ARG	G
G	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
	VAL	ALA	GLU	GLY	A
	VAL	ALA	GLU	GLY	G

(3) Mutations resulting in unstable polypeptide chains.
(4) The substitution of uncharged by uncharged, positive by positive and negative by negative charged amino acids.
(5) Lack of sensitivity of the electrophoretic method if one charged group is removed from a highly charged protein. This problem can be overcome by varying the *p*H of the measurements.

Harris (1975) provides a detailed survey of the enormous number of enzyme variants found in population surveys. He defines the term polymorphism or genetic polymorphism as follows: 'One may anticipate that many (mutants, alleles) will be very rare, but others will occur with an appreciable frequency, and indeed may be common enough to give rise to the well-known phenomenon usually referred to as genetic polymorphism'. In such cases the members of a population can be classified into two or more relatively common phenotypes due to their distinct alleles at certain loci.

The following different origins of polymorphism can be distinguished by genetic analysis.

(1) Multiple gene loci resulting from gene duplication and subsequent mutations can give rise to different forms of a protein in the same organism. This can also result in the formation of oligomeric proteins with different polypeptide chains in the same molecule, which has contributed to the evolution of functions as discussed below.
(2) Multiple allelism at a single locus is probably responsible for the majority of polymorphic forms of proteins.
(3) Post-translational modification of proteins causes the appearance of additional multiple forms of some proteins. This would not normally be regarded as genetic polymorphism, although the tendency could be genetically determined.

Approaching 100 allelic variants of the enzyme glucose 6-phosphate dehydrogenase are found in human populations. There appears to be no norm, no ideal enzyme with a unique amino-acid sequence. Of course, we only see those variants in which the correct active structure of the enzyme is preserved. With enzymes, just as with larger biological organisations, one could apply the criticism of Gould & Lewontin (1979) of what they call the Dr Pangloss syndrome. Voltaire makes Dr Pangloss in *Candide* expound the philosophy that this is the best of all possible worlds. The reader is referred to Gould (1979) for a wise and witty analysis of many misconceptions about evolution.

THE EVOLUTION OF OLIGOMERS (E.G. OXYGEN CARRIERS)

As mentioned above, gene duplication must have played an important role in the evolution of functions of oligomeric proteins. Many enzymes exist as aggregates of two, four and even six identical polypeptide chains, each with an independent active site. Many functions have been postulated for homogeneous oligomeric systems which are found, on careful kinetic analysis, to be artefactual (Gutfreund, 1975). Probably in some cases the early advantage of aggregation was increased stability. Subsequently interactions occurred between active sites across protomer interfaces and were preserved in those cases where cooperativity was advantageous for a particular key biological control mechanism.

Heterogeneous oligomers evolved either by enzyme units combining with unrelated regulatory units, which is not of interest in the present context, or by gene duplication and separate mutations. Of the latter phenomenon, the evolution of haemoglobin from single-chain oxygen carriers is the best-documented, if not the only, example. These proteins deserve some wider discussion than is possible here. Unfortunately we could not include a chapter devoted to them.

The large group of proteins which have a ferroporphyrin (protohaem) as a prosthetic group and act as oxygen carriers in a wide range of species have been of great interest to students of many different aspects of molecular biology. Their ubiquity, easy recognition and abundance have also contributed to the study of biochemical evolution and in turn biochemical genetics has contributed to our knowledge of their function. In vertebrates one distinguishes the tetrameric haemoglobins of the blood-stream from the monomeric myoglobins present in all forms of muscle. In other forms of animal life haemproteins occur in various oligomeric states in different tissues.

There are three distinct problems in the evolution of oxygen carriers. The evolution of tetrapyrroles (Hendry & Jones, 1980), the evolution of protoporphyrin-Fe^{2+}-globin (haemoglobin) (Ingram, 1963) and the evolution of iron or copper protein complexes. In principle it is easier to start a novel process with a protein than with another complex organic compound. Only one gene has to evolve to provide for the synthesis of a new protein while the synthesis of protohaem requires several enzymes and hence several genes. However, precursors of protohaem synthesis could have been useful for several purposes even before the formation of protoporphyrin with its ubiquitous use for plant and animal pigments in photosynthesis and respiration.

With the information available at the moment, one might hazard the guess that nature has several times solved the problem of designing oxygen carriers. There was no need for oxygen transport until photosynthesis produced enough oxygen for aerobic life. At that time there were likely to be plenty of tetrapyrroles around for the formation of chlorophyll, and cytochrome-like compounds. At the same time copper proteins (haemocyanins) and, to a lesser extent, non-haem iron proteins are used as oxygen carriers in many early classes of the animal kingdom. Although these metalloproteins have adapted to the varied needs of their surviving 'users' they have not acquired the enormous versatility of the haemproteins (Antonini & Brunori, 1971).

Haemoglobins are a very nice example of the modulation of the function of a prosthetic group by a great variety of minor changes of the protein. The wide range of oxygen binding affinities of the protohaem are controlled by the globin variants and their response to specific ions liganding at other sites. The large number of haemoglobin variants which occur in any population (Harris, 1975) have thus 'often?' proved useful in the adaptation of a subgroup to a new niche. The same phenomenon can, as mentioned above, help to elucidate the effect of changing certain amino-acid residues on the mechanism of oxygen binding and its control (see, for instance, Morimoto, Lehmann & Perutz, 1971).

RANDOM (GENETIC) DRIFT: NEUTRALISM

It has already been mentioned that other mechanisms, in addition to Darwinian natural selection, can contribute to evolution. There is no doubt that natural selection by competition is an important, perhaps the most important, cause of evolution. However, Darwin did not exclude other mechanisms or contributory factors. In recent years there have been very hot debates about the relative importance of different factors in the overall picture of the present genetic composition of populations. In particular, the establishment of neutral genes as a major factor in evolution has been the subject of an interesting controversy. It is not my purpose, and I do not feel competent, to discuss this subject in detail and to assess the relative importance of neutralism. However, since many of the experimental data used for this debate come from biochemical investigations, the principles involved should be mentioned here. Somewhat more detail will be found in the treatment of polymorphism by Harris (1975) and Smith (1975). A proper understanding

of the subject requires appreciation of its mathematical treatment by stochastic theory to which an introduction is found in Kimura & Ohta (1971).

In an infinite population the chances become negligible for a mutation to be passed on sufficiently to become established. The likelihood of a mutation becoming fixed is inversely proportional to the size of the population and directly proportional to the fecundity of its carrier. For detail and further references to the algebra involved in this reasoning the interested reader should consult Kimura & Ohta (1971). As a consequence one can postulate that in a large population even a mutation which can confer a considerable advantage on the carrier of its allele may not establish itself. It can also be argued that in a very small population a neutral or even a slightly disadvantageous mutation can establish itself due to relatively large fluctuations. The algebraic reasoning cannot be disputed and if neutral mutations arise in small populations, they will establish themselves. There remains, however, some doubt about the significant occurrence of neutral mutations and their fixation. Of course, the only test for this argument is whether some established variants are due to neutral mutations. On perusing the large amount of evidence offered, for instance by Harris (1975), one can well come to the conclusion that many different polymorphic forms of enzymes show no functional difference. Another argument for the neutrality of these polymorphs is the absence of variation in their distribution in different environments. It must be emphasised that many variants of enzymes have different functions within the same organism or in populations subjected to different environments.

Many investigations are in progress to analyse existing data and to design new experiments to test variants for their functional neutrality. It is probably fair to say that the case for or against is not a strong one and that the problem should stimulate more experimental work to provide data on mutation rates, genetic load etc. I shall conclude the discussion of this topic by mentioning two investigations which illustrate the present direction of some relevant biochemical thoughts. Place & Powers (1979) studied the catalytic efficiencies of variants of lactate dehydrogenase in the killifish in different environments. This study is typical of many designed to relate structural differences in enzymes which function at different temperatures and pressures. The authors conclude their paper with the statement: 'These catalytic differences between allozymes (another term for isozymes) . . . are consistent with the selectionist hypothesis'. So far, so good, but then nobody really denies that natural

selection exists as an important factor in evolution. The question is whether neutralism is also important.

It was mentioned above that the oligomeric structure of enzymes will contribute to their stability in solution. However, this stabilising effect will depend on specific contact areas and any mutation affecting the surface structure would be detrimental. Harris, Hopkinson & Edwards (1977) argue that the significantly lower incidence of polymorphism among multimeric compared to monomeric enzymes favours the neutralist hypothesis. It would be expected that more neutral mutations would occur with increasing oligomer size. This was not found to be the case in the investigations of Harris and his colleagues. In a recent paper by McConkey, Taylor & Phan (1979) there is a discussion of the bias obtained if only enzymes or only structural proteins are studied.

THE EVOLUTIONARY CLOCK

A topic which is likely to be of considerable interest, and which will require much further theoretical and experimental work for some time to come, is the study of the rate of evolution. The methods described by Peacock in chapter 3 can be extended and combined with the dating of particular branch points of the phylogenetic trees to construct what has been called an evolutionary clock. Several questions arise which are stimulating valuable investigations. In general, biologists are divided between those who believe that there have been periods of considerable evolutionary activity, followed by comparatively quiescent times, and others who consider it likely that there were only small fluctuations on a steady rate. An important contribution to these studies is the study of the rates of mutations. For this purpose, as for the construction of phylogenetic trees, complete polypeptide or nucleotide sequence information is most desirable. If the number of mutations is to be determined in a particular protein, say at different stages of primate evolution, it is of course much less work to determine the changes in sequence after the first complete analysis. Methods such as 'finger printing', first used by Ingram (1963) to define the single mutation to sickle-cell haemoglobin, select the small peptides which contain the changes. Inspection of the factors quoted earlier as contributions to the under-estimation of the number of mutations will show that only some of these are eliminated by complete sequence analysis of the gene products – polypeptide chains. Recent dramatic improvement in methods for DNA sequencing makes it likely that in future much more information about rates of evolution will be