JACQUES NINIO
ROBERT LANG

Molecular Approaches to Evolution

PRINCETON LEGACY LIBRARY

Molecular approaches to evolution

Molecular approaches to evolution

Jacques Ninio

National Centre
for Scientific Research, Paris

Translated by
Robert Lang
National Institute for Medical Research,
Mill Hill, London

Princeton University Press
Princeton, New Jersey

First published in the United State of America 1983 by
Princeton University Press,
41 William Street, Princeton, New Jersey 08540

Printed in Great Britain

Library of Congress Cataloging in Publication Data
Ninio, J.
Molecular approaches to evolution.
Translation of: Approches moléculaires de l'évolution.
Bibliography: p.
I. Chemical evolution. I. Title.
QH371.N5613 1982b 577 82–3766
ISBN 0–691–08313–4 AACR2
ISBN 0–691–08314–2 (pbk.)

Translated by Robert Lang from *Approches moléculaires de l'évolution*

First English edition published by Pitman Books Ltd, 1982

Contents

1 Clearing the ground

We can understand evolution without really knowing what life itself is! This view, held by the great biologist Haldane, was characteristic of a half century of evolutionism which produced models and calculations in the absence of any intimate knowledge of cellular logic. The evolutionist used to try to calculate the outcome of battles for survival just as a General evaluates the chances of victory on a battlefield: will one side score a resounding victory, will the two forces neutralize each other or will they occupy the field alternately? The theory of evolution by natural selection lends itself to all sorts of ramifications. Its principal theme is that since many more individuals of each species are born than can reach reproductive age, those best fitted for the struggle for existence will have more chances of leaving descendants and the characters they carry, passed from generation to generation, will eventually become fixed within the species. If new and even fitter variants appear, they will gradually eliminate the former ones to their own advantage.

I support the doctrine of natural selection in the same way that I recognize the truth of the syllogism that Socrates is a man, all men are mortal, therefore Socrates is mortal – but I go no further than that. A correct idea can become sterile. Renaissance scholars believed that the study of syllogism would enable them to work out the whole of logic. Instead they merely got bogged down in an exploration of Baralipton, Darapti and 254 other outdated forms of syllogism. Of course, evolutionists had to examine closely the fundamental propositions of Wallace and Darwin and assess the possible influences of all imaginable factors: small or large populations, isolated or panmictic, sexed or not. But today we must recognize that this area of evolutionary biology tends to become confined to the study of infinitely increasing variants.

In renewing our conception of living things, molecular biology sheds new light on an evolutionism which until now has been too formal and helps us to redefine the problems. At first molecular biologists were content to verify at their level the validity of concepts developed by the evolutionists 30 years earlier. Duplication of chromosome segments became duplication of DNA segments; comparative anatomy was replaced by comparative molecular anatomy. Fortunately, things did not rest there. After this consolidation phase a new generation of work began which posed the problem of evolution in a completely new manner. In the former view an animal predator with good eyes would be able to capture its prey more easily than one with

defective vision and would therefore be selected for. But what made possible the existence of a visual apparatus in animals? Why was this solution realizable and not another? Or, if other solutions were available, what were they? In summary, our aim now is to tackle a problem which not so long ago seemed insoluble: how to evaluate, having taken into account the state of organization of life at any given moment, the various possibilities which are apt to manifest themselves. This is the primary aspect of the new set of problems posed by studying evolution. A secondary aspect is how to decide, once a novelty has appeared, if it will be established or rapidly eliminated. I shall therefore discuss evolution, as far as one can today, in the new perspective as applied to the molecular logic of living things. The genetic code is at the heart of this logic. Above all, we want to know why the code exists and through what tentative stages cellular organization, now so well formed, could have passed. We will tackle step by step all the interesting concepts in molecular biology which revolve around the genetic code.

At first sight the layout and order of the contents will appear disconcerting, but nothing has been left to chance. Those who like to have everything neatly cut and dried will be disappointed. I am not presenting a standard product, aseptic and reassuring, to be offered in a supermarket, but a discourse which sticks to the facts of current research, which refuses to gloss over contradictory aspects of present ideas and which allows, albeit tentatively, social and ideological connotations to show through. Instead of following each statement with ten lines of support, ten against and three of compromise, I concentrate on only one aspect and further on, when I show how the problems can be posed in other ways, I discuss the neglected aspects in other contexts, which lend them their true importance. A major aim of the layout which I have adopted is to produce a text which can be read with profit and interest by readers at very different levels. Technical terms have been kept to a minimum. To avoid tedious exposition of details of molecular biology I have dispersed reminders throughout the text and have placed them in an evolutionary context as soon as they appear.

Initially, of course, the evolutionary approach will be rudimentary: the kind in which molecules from different organisms are compared to deduce ancestral relationships between species. This aspect of molecular evolution has already received wide coverage (it is discussed in Chapter 3). On the other hand, everything else is practically new. No other textbook on evolution deals with comparisons in three dimensions, with acquisitive evolution and with the stability of the genetic code. No other work since the remarkable books of Woese (1967) and Yčas (1969) has discussed the genetic code in depth. After briefly describing constituents (proteins and nucleic acids) and comparing them we shall move on to mechanisms, to the most important cellular processes: information transfer, replication and genetic translation. Having revealed the logic behind these mechanisms, we shall attempt to understand how they could have emerged. Lastly, we shall reach a third

stage. The cell is not simply a collection of molecules or a juxtaposition of mechanisms but a coherent entity in which each part is subject to constraints of compatibility with the rest. We shall attempt to clarify this to form concepts which will enable us to understand an evolution which proceeds from one coherent whole to another which is equally coherent.

Free to annoy the purist, I have often preferred to use simple language, close to the spoken word, which is very effective for putting ideas across. The research worker in the laboratory, when he is not writing administrative reports which will be judged by the quality of the typing uses everyday words and images borrowed from daily life. This can hardly be denied by the evolutionists, who learnt selection from cattle breeders and took their folklore of family trees from the genealogists. In the *Origin of Species*, Darwin wrote in an accessible style about ideas which were profoundly original for the scientific mentality of his time. Nevertheless, on one point my attitude is different from that of the Master. The progress of ideas is to a large extent a collective endeavour. Although I have contributed to the modern conception of the genetic code, I do not claim any exclusiveness but place my work firmly in the mainstream of modern evolutionary theory.

I thank all those who have helped me in this enterprise, especially the five colleagues and friends who went through the first draft with a fine-toothed comb: Jean-Pierre Dumas, Jerôme Lavergne, Jeanine Rondest, Pierre Roubaud and Michel Volovitch and those who helped me to review the English version: Margaret and Richard Buckingham, Dick d'Ari and Liliane Assairi. I also thank all those who, through the interest which they showed for the subject, encouraged me to set down these reflections on evolution. Above all, my gratitude to Robert Lang: I recognize my child in his translation. Luck or prescience, the French text, written in 1977–1978 needed little modification to cope with the latest advances. The Bibliography was updated in July 1980, then in October 1982, and adapted for an English-speaking audience.

2 The chemistry of life

At any given moment thousands of chemical reactions are occuring in the living cell. The food absorbed serves both as fuel, which supplies energy, and as material for constructing new structures or for replacing parts used by the cell. Only plants and photosynthetic bacteria can draw their energy directly from sunlight. The transformations, degradations and syntheses are chemical in nature. Each reaction in the cell is carried out and controlled by a molecule which is designed for this purpose – an enzyme. The enzyme interacts with the substances which have to be chemically combined and accelerates the speed of the reaction much more effectively than any mineral or organic catalyst known to chemists. The general biochemistry by which the cell synthesizes enzymes is common to all terrestrial life forms with only minor variations from the most rudimentary to the most highly evolved creatures.

If reactions were left to occur naturally in water, a molecule A would be able to change into many different molecules: B, C, D, etc. D in its turn would result in L, M, O, P, etc. The enzyme, by accelerating only some of these transformations, orientates the processes along very special routes. Everything done chemically in living creatures is feasible without enzymes. But their presence, by selectively accelerating certain reactions, makes the chemical machine run and keeps it stable in its motion. In chains of reactions in the cell, where the product of each reaction is the substrate of the following one, all intermediates are maintained at workable concentrations. There are neither flat spots nor bursts of power. The speeds of the reactions in a chain and the concentrations of the intermediates are well adjusted to each other. The cellular machinery runs smoothly despite environmental variations.

Another feature of chemical transformations in the cell, which is linked to the first, is that they occur gradually. Reactions which would produce or consume too much energy are split up into a large number of steps. For example, under cellular conditions, the breakdown of glucose to water and carbon dioxide must release about 2.85 MJ/mol (680 kcal/mol). Cellular reactions use energies in the order of about 42 kJ/mol (10 kcal/mol), often less. Reactions which consume energy are coupled to others which produce it; one way of raising a weight is to link it to a counterweight by a pulley or winch and to lower the counterweight. By carefully balancing the pulley it is possible to control better the raising or lowering of the weight. In cellular reactions the counterweight principle is provided by the decomposition of ATP (we shall

see later what this molecule is like) to ADP or AMP. The energy released is in the order of 50 kJ/mol (12 kcal/mol). In passing we should mention the existence of a second system based on equilibria of oxido-reduction (the $NADH_2$/NADH system). Glucose, converted into 'energetic small change', provides about 38 molecules of ATP, the universal counterweight. But ATP is not in itself an 'energy capsule'. Every chemical reaction is in principle reversible. If we start with compounds A and B and end up with C and D according to $A + B \rightarrow C + D$, then by mixing C and D at high enough concentrations we must be able to form A and B thus: $C + D \rightarrow A + B$. One can recover energy by breaking down a molecular species – glucose or ATP here – only when the species is in excess of the calculated equilibrium concentration. 'Glucose energy' is not converted into 'ATP energy', but a glucose/CO_2 imbalance is converted into an ATP/ADP imbalance.

The enzyme, like every catalyst, accelerates both forward and back reactions by the same amount so that the concentrations of the reactants tend towards their equilibrium values. If the cell has to convert A into C, then instead of using the natural reaction $A + B \rightarrow C + D$, it can substitute for B a molecule analagous to ATP. The thermodynamic constraints are no longer the same and a better C/A ratio can be obtained. In another phase of the cell cycle it might be advantageous to move the reaction in the opposite direction, from C to A. The old enzyme is put out of service and a new one comes into action which links the energy counterweight in the opposite direction. Regulatory mechanisms allow enzyme activities to be modulated as required. If a product is in excess its formation is usually automatically slowed down, since it inhibits one or more of the enzymes in the reaction sequence which leads to its synthesis. There exists another, more elaborate form of regulation, discovered by Jacob and Monod. The presence of a substrate A produces a signal which switches on the production of an enzyme capable of transforming A. In the absence of A, the synthesis of this enzyme is repressed. Thus, superimposed on the network of chemical reactions in the cell, there is a still more complicated system of regulatory actions, interactions and retroactions which adjust the quantities and activities of each enzyme as functions of the concentrations of cellular compounds.

Enzymes and other cellular macromolecules tend to decompose when in contact with water. The small molecules used by the cell are usually relatively unoxidized. When they are exposed to oxygen in the air they are eventually converted into substances which are of no use for cells. Thus substances which form spontaneously in the terrestrial environment bear little resemblance to biological substances. The great idea of the Soviet scientist Oparin, formulated in 1924, was that the chemical environment which prevailed at the Earth's surface when it was formed could have been very different from that of today. Later Haldane introduced the concept of a primitive terrestrial atmosphere, low in oxygen but rich in reducing gases such as methane and ammonia. Oparin specified major ways in which biologically important

molecules could have been formed spontaneously under these primitive conditions.

A quarter of a century later, a young American student, Stanley Miller, constructed a chemical reactor which simulated primitive conditions on Earth. He produced electrical discharges in a flask containing methane and ammonia. Products were formed which were carried away in steam and continued to react in a second flask, which took the place of the primitive ocean. On the seventh day Miller saw that the results were good and stopped the experiment. Deposits had formed, turning the water brown. Miller analysed these after acid degradation and presented his supervisor with results about which he had no need to blush. Several biological substances were detected; notably two amino acids, glycine and alanine. Aspartic and glutamic acids were identified later. Thus Miller in 1953 supplied the first experimental support for the theory of Oparin and Haldane. I had the privilege of meeting Oparin while he was in semi-retirement, at the Bakh Institute, surrounded by venerating collaborators. I found in him an intellectual vigour and a wide-ranging and penetrating vision which are not often met in active scientists. Why had no-one around him in 30 years attempted to perform a Miller's type experiment? He saw this as unnecessary since Miller's experiment was fundamentally equivalent to that of Wohler dating from 1832. And in fact several bridges between the chemistry of life and inorganic chemistry had been established over the previous century. This in no way detracts from Miller's achievement, since his experiment was above all a beginning on which a new branch of science, prebiotic chemistry, was going to be built.

Having explored possible ways of synthesizing amino acids, prebiotic chemistry has succeded in proposing plausible mechanisms for the synthesis of sugars, nucleotide bases, phosphorylated compounds and coenzymes – all the small molecules which play important roles in living cells. These syntheses apparently include the preliminary formation of some highly reactive compounds such as formaldehyde (HCHO) and hydrogen cyanide (HCN). Astronomers, with their radiotelescopes, can detect these compounds in space. A notable result of the last 10 years is that a large number of reactive radicals, whose importance has been emphasized in prebiotic chemistry, have been found in interstellar space in surprisingly high concentrations.

Can we find traces of synthetic pathways of prebiotic chemistry in present-day metabolism? Let us take as an example a metabolic pathway leading to the synthesis of an amino acid X:

$$U \xrightarrow{E_1} V \xrightarrow{E_2} W \xrightarrow{E_3} X$$

Suppose that the enzyme E_1 catalyses the reaction U→V, that E_2 catalyses V→W and E_3 catalyses W→X. According to an ingenious argument of

Horowitz's it is reasonable that these three enzymes might have arisen in the course of evolution in the reverse order from that which they have in the scheme. Imagine some very primitive organisms which used X to make their proteins. In those distant times, X was abundant, produced by natural reactions at the Earth's surface, so the cell had no need to synthesize it. Little by little, the reserves of X became exhausted, partly because cells used it up and partly because it decomposed spontaneously according to the sequence: X→W→V→U. Then the easiest way of getting X was to synthesize it from its first decomposition product W, hence enzyme E_3 which catalyses W→X. When W was exhausted in its turn, the appearance of an enzyme E_2 which reformed W from its decomposition product V, together with E_3 which was already present, allowed the lack of X to be made good. The idea we may glean from this hypothesis is that present-day metabolic pathways go in the opposite direction to primitive ones. Other authors, such as René Buvet, see present-day metabolism as a transposition of the great pathways of primitive chemistry. Since the cell often breaks down biological compounds produced by another organism to synthesize closely similar compounds, both points of view have something in their favour. But in the bioenergetic view of the beginning it appears that metabolic pathways obey a third logic: instead of direct and abrupt synthetic routes the cell would prefer gently sloping, winding paths which would allow energetic differences to be split up.

Enzymes are synthesized in the cell by arranging simpler chemical units, amino acids, in a certain order and linking them end to end. Twenty amino acids are used in the synthesis of enzymes and other proteins. Each protein consists of an exact bonding of 50, 100, 1000 or more amino acids, which follow each other in a determined order. Thus human lysozyme, which makes tears effective as a bactericide, starts with the amino acid series: lysine, valine, phenylalanine, glutamic acid, arginine . . . and this sequence is defined until the end. Naturally, other proteins have different amino acid sequences. The chemical procedure by which the cell synthesizes its proteins is extraordinarily precise, as will be shown later. Changing a single amino acid in a protein is sometimes sufficient to change its properties. People with sickle-cell anaemia have an 'abnormal' haemoglobin with a valine instead of a glutamic acid at position 6 in a chain of 146 amino acids. The name enzyme is reserved for proteins with catalytic activity. Proteins also have other roles. Regulatory proteins switch enzyme activity on or off. Proteins are one of the basic materials of all biological tissues: muscle, cartilage, hair etc. In membranes they are in charge of the cells' frontier posts through which products are pumped into the surroundings or excreted. The antibodies (Chapter 14), which protect us against biological aggression, are proteins.

The amino acids, except proline, have the general formula: