ON THE EVOLUTION OF BIOCHEMICAL SYNTHESES

BY N. H. HOROWITZ

SCHOOL OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY, CALIF.

Communicated April 23, 1945

Although it has been recognized for a long time that the biochemistry of the organism is conditioned by its genetic constitution, a more precise definition of this dependence has not been possible until recently. A considerable amount of evidence now exists for the view that there is a one-to-one correspondence between genes and biochemical reactions. This concept, foreshadowed in the work of Garrod[1] on human alcaptonuria, accounts in a satisfactory way for the inheritance of pigment formation in guinea pigs,[2] insects[3] and flowers,[4] and the synthesis of essential growth factors in *Neurospora*.[5] It appears from these studies that each synthesis is controlled by a set of non-allelic genes, each gene governing a different step in the synthesis. As to the nature of this control, it is probable that the primary action of the gene is concerned with enzyme production. That genes can direct the specificities of proteins has been shown in the case of many antigens,[6] while several mutations demonstrably affecting the production of enzymes have been reported.[6] Evidence on the postulated gene-enzyme relationship is in most cases, however, still circumstantial; this is partly because of technical difficulties involved in the study of synthetic, or free-energy consuming reactions *in vitro*, and partly because of the insufficiency of biochemical information on those reactions which happen to be susceptible of genetic analysis.

As a corollary of the above hypothesis, each biosynthesis depends on the direct participation of a number of genes equal to the number of different, enzymatically catalyzed steps in the reaction chain. In attempting to account for the evolutionary development of such a reaction chain one meets in a clear form the problem of explaining macroevolutionary changes in terms of microevolutionary steps. The individual reactions making up the chain are of value to the organism only when considered collectively and in view of the ultimate product. Regarded individually, intermediate substances cannot, in general, be assumed to have physio-

logical significance, and the ability to produce them does not of itself confer a selective advantage. An example from *Neurospora* genetics will serve to illustrate this point. At the present time seven different genes are known to be concerned in the synthesis of arginine by the mold.[7] The inactivation of any one prevents the synthesis from taking place. On the basis of the above hypothesis, at least seven different catalyzed steps must occur in the synthesis. Several of the steps have been identified and controlling genes assigned to each. Two of the intermediates in the chain have been shown to be the amino acids ornithine and citrulline. Unlike arginine, neither of these substances is a general constituent of proteins. Aside from their function as precursors, they are apparently of no further use to the organism.

While the above example probably represents the general case, there are also well-known instances in which precursors serve independent functions. Thus, arginine, glycine and methionine are precursors of creatine in the rat,[8] but the synthesis goes through the non-functional intermediate, glycocyamine. On the other hand, acetylcholine may be synthesized from choline in one step.[9] In cases such as these, the problem is that of accounting for the synthesis of the precursors.

Since natural selection cannot preserve non-functional characters, the most obvious implication of the facts would seem to be that a stepwise evolution of biosyntheses, by the selection of a single gene mutation at a time, is impossible. It will be shown below that this is not a necessary conclusion, but that under special conditions the stepwise evolution of long-chain syntheses may occur. First, however, an alternative to stepwise evolution will be considered; that is, the origin of a new reaction chain through the chance combination of the necessary genes.

Although the probability of the origin of a useful character through the chance association of many genes may be small, it is never zero. Indeed, a consideration of the statistical consequences of the interaction of mutation, Mendelian inheritance, and natural selection has led Wright[10] to the conclusion that such chance associations may be of major importance in evolution. He has analyzed the evolutionary possibilities of various types of breeding structures and has shown that under certain conditions an extensive trial and error mechanism exists, whereby the species can test numerous combinations of non-adaptive genes. The breeding structure which most favors this type of evolution is that of a large population divided into many small, partially isolated groups. Within each group the cumulative effects of the accidents of sampling among the gametes are of major significance in determining gene frequencies, but the penalty of fixation of deleterious genes, ordinarily incurred under inbreeding, is avoided by exchange of migrants with other groups. The pressures of forward and reverse mutations, which between them determine an equilib-

rium frequency for non-adaptive genes in large, random-breeding populations, become of minor importance. As a consequence, a random drift of gene frequencies occurs. If, by chance, one group finds a particularly favorable combination of genes, a process of intergroup selection comes into play, whereby the favorable combination is spread to the population at large.

This model provides a means for the evolution of a new gene combination in spite of unfavorable mutation rates to active alleles and in the absence of selection of individual genes. It is thus favorable for the evolution of systems of individually non-adaptive, but collectively adaptive, genes. The effectiveness of the process would seem to be strongly dependent on the size of the gene combination required, however, decreasing approximately exponentially with increasing numbers of genes, other factors remaining constant. There would result a tendency toward the evolution of short reaction chains involving the recombination of molecular units already available. There is no doubt that a conservative tendency of this sort actually exists in nature. The wide variety of biologically important compounds built up on the pyrrole nucleus, to mention but one example, is a case in point.

The application of Wright's theory to the particular problem under consideration is limited by the fact that it operates only under biparental reproduction. It is probable that a large number of basic syntheses evolved prior to sexual reproduction. The universal distribution among living forms of certain classes of compounds—viz., the amino acids, nucleotides and probably the B vitamins—identifies them as essential ingredients of living matter. The synthesis of these substances must have evolved very early in geologic time, as a necessary condition for further progress, although loss of certain syntheses may have occurred in the later differentiation of some forms. It is therefore desirable to search for another solution of the problem applicable to compounds of this type, preferably one in which a minimum burden is placed on chance and a maximum one on directed evolutionary forces. It is thought that the following suggestion, while definitely a speculation, offers a possible solution along these lines.

In essence, the proposed hypothesis states that the evolution of the basic syntheses proceeded in a stepwise manner, involving one mutation at a time, but that the order of attainment of individual steps has been in the reverse direction from that in which the synthesis proceeds—i.e., the last step in the chain was the first to be acquired in the course of evolution, the penultimate step next, and so on. This process requires for its operation a special kind of chemical environment; namely, one in which end-products and potential intermediates are available. Postponing for the moment the question of how such an environment originated, consider the

operation of the proposed mechanism. The species is at the outset assumed to be heterotrophic for an essential organic molecule, A. It obtains the substance from an environment which contains, in addition to A, the substances B and C, capable of reacting in the presence of a catalyst (enzyme) to give a molecule of A. As a result of biological activity, the amount of available A is depleted to a point where it limits the further growth of the species. At this point, a marked selective advantage will be enjoyed by mutants which are able to carry out the reaction $B + C = A$. As the external supplies of A are further reduced, the mutant strain will gain a still greater selective advantage, until it eventually displaces the parent strain from the population. In the A-free environment a back mutation to the original stock will be lethal, so we have at the same time a theory of lethal genes. The majority of biochemical mutations in *Neurospora* are lethals of this type.

In time, B may become limiting for the species, necessitating its synthesis from other substances, D and E; the population will then shift to one characterized by the genotype $(D + E = B, B + C = A)$. Given a sufficiently complex environment and a proportionately variable germ plasm, long reaction chains can be built up in this way. In the event that B and C become limiting more or less simultaneously, another possibility is opened. Under these circumstances symbiotic associations of the type $(F + G \neq C, D + E = B)(F + G = C, D + E \neq B)$ will have adaptive value.

This model is thus seen to have potentialities for the rapid evolution of long chain syntheses in response to changes in the environment. As has been pointed out by Oparin[11] the hypothesis of a complex chemical environment is a necessary corollary of the concept of the origin of life through chemical means. The essential point of the argument is that it is inconceivable that a self-reproducing unit of the order of complexity of a nucleoprotein could have originated by the chance combination of inorganic molecules. Rather, a period of evolution of organic substances of ever-increasing degree of complexity must have intervened before such an event became a practical, as distinguished from a mathematical, probability. Or, put in another way, any random process which can have produced a nucleoprotein must at the same time have led to the production of a profusion of simpler structures. Oparin has considered in some detail the possible modes of origin of organic compounds from inorganic material and cites a number of known reactions of this type, together with evidences of their large-scale occurrence on the earth in past geologic ages. He concludes that in the absence of living organisms to destroy them highly complex organic systems can have developed. The first self-duplicating nucleoprotein originated as a step in this process of chemical evolution. The origin of living matter by physicochemical means thus

presupposes the existence of a highly complex chemical environment.

To summarize, the hypothesis presented here suggests that the first living entity was a completely heterotropic unit, reproducing itself at the expense of prefabricated organic molecules in its environment. A depletion of the environment resulted until a point was reached where the supply of specific substrates limited further multiplication. By a process of mutation a means was eventually discovered for utilizing other available substances. With this event the evolution of biosyntheses began. The conditions necessary for the operation of the mechanism ceased to exist with the ultimate destruction of the organic environment. Further evolution was probably based on the chance combination of genes, resulting to a large extent in the development of short reaction chains utilizing substances whose synthesis had been previously acquired.

[1] Garrod, A. E., *Inborn Errors of Metabolism*, Oxford University Press (1923).

[2] Wright, S., *Biol. Symposia*, 6, 337–355 (1942).

[3] Ephrussi, B., *Quart. Rev. Biol.*, 17, 327–338 (1942).

[4] Lawrence, W. J. C., and Price, J. R., *Biol. Rev.*, 15, 35–58 (1940).

[5] Horowitz, N. H., Bonner, David, Mitchell, H. K., Tatum, E. L., and Beadle, G. W., *Am. Nat.*, in press (1945).

[6] Summarized in Wright, S., *Physiol. Rev.*, 21, 487–527 (1941).

[7] Srb, A., and Horowitz, N. H., *Jour. Biol. Chem.*, 154, 129–139 (1944).

[8] Summarized in Schoenheimer, R., *The Dynamic State of Body Constituents*, Harvard University Press (1942).

[9] Lipmann, F., *Advances in Enzymology*, 1, 99–162 (1941).

[10] Wright, S., *Bull. Am. Math. Soc.*, 48, 223–246 (1942). Contains summary of earlier papers.

[11] Oparin, A. I., *The Origin of Life*, trans. by S. Morgulis, Macmillan, New York (1938).

STRAIN SPECIFICITY AND PRODUCTION OF ANTIBIOTIC SUBSTANCES. V. STRAIN RESISTANCE OF BACTERIA TO ANTIBIOTIC SUBSTANCES, ESPECIALLY TO STREPTOMYCIN*

BY SELMAN A. WAKSMAN, H. CHRISTINE REILLY AND ALBERT SCHATZ

NEW JERSEY AGRICULTURAL EXPERIMENT STATION, RUTGERS UNIVERSITY,
NEW BRUNSWICK, NEW JERSEY

Different strains of the same species of bacteria are found to vary greatly in their sensitivity to a given antibiotic substance. This phenomenon has an important bearing upon the utilization of the substance for chemotherapeutic purposes, where a knowledge of the sensitivity of the particular strain of a given organism responsible for a certain disease becomes of paramount importance.