18 December 1964, Volume 146, Number 3651 SCIENCE

# Organisms and Molecules in Evolution

Studies of evolution at the molecular level lead to greater understanding and a balancing of viewpoints.

George Gaylord Simpson

It is universally recognized that molecules of biological importance may evolve—that is, they may change in the course of time as have the organisms in which they occur. Some molecules, like adenosine triphosphate, are so nearly universal and invariable as to suggest no evolutionary sequence, but many others surely have evolved, notably groups of proteins and, obviously, DNA. Before the importance of DNA was known, Florkin (*1*) had already discussed the systematics and evolution of various families of molecules. In such instances evolutionary interpretation of the biochemists' findings requires information from paleontologists and systematists, information especially on the time scale involved and the phylogeny and relationships of the species in which varying molecules are to be compared. An example is the hypothesis that serum proteins (*2*) or cytochromes (*3*) have changed in a regular if not linear manner with respect to time—that they have evolved by some sort of internal constant-rate mutational process and not in an irregular or a specifically adaptive way. In fact, when the data are replotted with what seem to be the most probable time coordinates they indicate that the hypothesis is incorrect or, at least, that these data do not support it. Williams now tells me that the hypothesis has been modified, but it exemplifies the clarifying confrontation of molecular and organismal data.

Other interesting examples of such confrontation arise from further studies of serum proteins, such as that by Goodman (*4*). Phylogenetic relationships of the animals concerned, primates in this case, are inferred from the apparent degrees of homology in their various serum proteins. The lineages thus inferred then permit conclusions as to the evolution of the proteins themselves. Similar inferential methods have been applied to the evolution of hemoglobins, also in primates, by Hill and the Buettner-Janusches (*5*). When phylogeny is inferred from the molecular data and molecular evolution is inferred in turn from that phylogeny, there is an element of circularity, which does not wholly invalidate the method but does warrant some reservations. A necessary cross-check is to arrange the molecular data in the framework of a phylogeny based entirely on nonmolecular evidence. It should be mentioned in passing that this, too, has sometimes led to semi-circular reasoning when molecule-based phylogeny has been compared with phylogeny with other bases: agreement between the two has been taken as the requisite validation of the molecular approach to phylogeny, but nonagreement has been taken as evidence of the greater reliability of the molecular method.

However, the most important reason for relating organismal and molecular evolution to each other is not simply the testing of hypotheses or the validation of methods. It is the balancing of points of view and the achievement of more complete explanations. Wald (*6*) has said that "living organisms are the greatly magnified expressions of the molecules that compose them." Anfinsen (*7*) believes that "we may almost define the life sciences as those concerned with the elucidation of the mechanisms by which molecules exert their specific actions in living cells." In fact there are many respectable and even eminent students of the life sciences who have no concern whatever with molecules or their actions. Concentration on one level of organization to the practical exclusion of others is often a necessity of specialized research, but nowadays almost everyone agrees that eventual understanding of relationships between levels is also necessary. Sonneborn (*8*) has emphasized the fact that molecular genetics could only have arisen through, and would now have little meaning apart from, "classical" or Mendelian organismal genetics. Weiss (*9*) has pointed out that there is a "cellular control of molecular activities" as well as a molecular control of cellular activities. There is also an organismal control of cellular activities, and, for that matter, a populational control of organismal activities. Indeed both Wald and Anfinsen, in the works from which one-sided aphorisms have been quoted, were concerned with relationships of molecules to higher organizational levels in evolution.

The sort of problem that can arise from a limited approach is exemplified in a recent article by Mora (*10*). He points out that living organisms have

---

Dr. Simpson is Alexander Agassiz Professor of Vertebrate Paleontology at the Museum of Comparative Zoology, Harvard University, Cambridge, Mass. This article is adapted from a paper prepared for a colloquium on protides of the biological fluids, in Bruges, Belgium, in the summer of 1964.

a teleological or purposive aspect which he proposes to label "urge." He finds that this aspect is inexplicable at the molecular level as hitherto studied. He proposes, but does not describe, a new approach, to be frankly permeated by teleology. Although he seems to think or hope that this may still be naturalistic, he does not clearly state what a naturalistic teleology might be. Now, this is precisely the problem with which organismal biologists have been coping for generations. Unknown, it would seem, to some biochemists, they have achieved a naturalistic (or, in a sense, materialistic) explanation of what is now often called [after Pittendrigh (*11*)] the teleonomic aspect of organisms. The teleonomic, or *apparently* teleological or purposive, characteristics of organisms are adaptations. They include "urge" itself in Mora's sense, its manifestations, and its results in the activities of individuals and the evolution of populations. Teleonomic adaptations arise in the course of evolution, and the factor governing their origin and maintenance is natural selection. That is surely as true at the molecular level as at any other. However, the ramifications of natural selection at various levels are far from simple.

## Natural Selection

The process of natural selection, as now understood, is complex rather in its concrete working and its interactions than in its basis. That basis is simply differential reproduction correlated with genotypic constitution. If some individuals in a population have more surviving and breeding offspring than others, and if there is a consistent average difference, however small, in the genotypes of those who have more and those who have fewer, that is natural selection at work. The actual selection—that is, the determination of which individuals have more or fewer offspring that survive to breed in their turn—is an interaction between environment, in the broadest sense, and the population, in all its individuals throughout their complete ontogenies. Aspects of this process are discussed at length in recent works (see, for example, *12–14*) which supply many details not given here.

Natural selection requires, first, reproduction and, second, hereditary variation of such a kind as to influence the success of reproduction under existing circumstances. When those factors are present, natural selection *necessarily* occurs. In precellular evolution [a principal concern for Mora (*10*)] it necessarily began when there were replicating molecules that differed in the rate or efficiency of replication (see *15*). However, the pertinent unit is not the replicating molecule but the reproducing system. This was presumably a molecule at first but became a cell at the protistan level, and is a dynamic unicellular-to-multicellular ontogenetic individual at metaphytic and metazoan levels. Selection acts on the whole phenotype and can single out genes only to the extent that they have phenotypic effects separable both phenotypically and genetically from those of other genes. Although selection apparently does act in an analytically separable way on some particular molecules, it evidently does not do so as a rule. It usually acts on supramolecular phenotypic characters, on whole complexes of them, or indeed on all of them at once. Since most genes are pleiotropic and most characters are polygenic, it follows that selection usually is not concentrated on single genes, as might appear from the necessarily oversimplified models first formulated by population geneticists. Although the connection is not yet well understood, this presumably means also that it is unusual (it may even be impossible) for intermediary molecules such as enzymes and other proteins to be selected for or selected against independently of other molecules.

## Effect of Selection on Particular Features

In considering the effect of selection on particular features of an organism, it is important to judge how far these are in one direction from the genes and in the other direction from the phenotypic characters directly subject to selection. Behavior is subject to particularly strong selection, and it is probably farthest removed from the genes and also most elaborately polygenic as a rule. Some single-gene determinants of behavior are known, but they are exceptional (see *16*). Proteins or, at least, intracellular enzymes are believed to be almost directly and uniquely determined by one or a few particular genes. The effect of selection will surely be influenced by the length of the functional chain from the genes to the character selected for or against. As a rule, with exceptions, the effect becomes more, not less, diffuse and less, not more, direct as the level of the gene is approached.

Zuckerkandl (*17*) has argued that a molecule like hemoglobin is preferable to most "structural," or more remotely phenotypic, characters for the determination of affinities because it is so near the genes, so nearly a direct reflection of part of the DNA code. It may be added that hemoglobin is so literally vital that natural selection may here act at a level near the gene. Those are advantages in certain respects, but they are accompanied by disadvantages, and the more distantly phenotypic approach also has advantages, as Zuckerkandl notes but possibly understresses. Zuckerkandl has shown that, "from the point of view of hemoglobin structure, it appears that gorilla is just an abnormal human, or man an abnormal gorilla, and the two species form actually one continuous population." From any point of view other than that properly specified, that is of course nonsense. What the comparison seems really to indicate is that in this case, at least, hemoglobin is a bad choice and has nothing to tell us about affinities, or indeed tells us a lie. (It does show that men and gorillas are rather closely related, but that has long and more accurately been known from traditional morphological comparisons.) Of course, as Zuckerkandl points out, we should use not just one kind of molecule but many, preferably proteins. However, if one can be misleading, so can many! (Let me add that Zuckerkandl's discussion of the phylogenetic interpretation of molecular data is invaluable and, unfortunately, almost unique.)

In some respects it is a drawback that hemoglobin, various enzymes, and some other proteins are so near to the genes in the functional chain. It means that each sample is genetically determined by, and therefore provides a sample of, only an extremely minute part of the whole genetic system—apparently only two genes in the case of hemoglobin and probably only one for many enzymes. The farther a character is from the genes, the more likely it is to sample a number of genes or a really significant part of the whole genetic system. The complexity of the genetic determination of a characteristic is a positive advantage, not a disadvantage, when the purpose is to determine affinities of whole organisms. Moreover, such characters are in almost all cases those which were in fact subject to selection. On an average, the farther we are from genes the nearer we are to the action of selection, and thus the better able we are to interpret the adaptive processes involved.

When, as is usual, selection is on the phenotype and well removed from the genotype, all that matters is that the genotype should in fact result in the selectively favored phenotype under the existing conditions of development. In this sense, or beyond that point, it really can be said that the genotype does not matter in adaptive evolution. There is ample evidence (much of it summed up in *14*, with references) that genotype-phenotype determination is not unique in either direction. Phenotypes that are apparently identical and that seem to be equal in the face of selection can have markedly different genotypes. There are also many systems—genetic, ontogenetic, and selectional—that tend to channel phenotypic development in the face of considerable change or variation in genes and hence, presumably, also in many families of macromolecules (*18*). I am arguing not that any one kind of evidence on evolution—genetic, molecular, phenotypic, or other—is superior but, on the contrary, that no one kind suffices in itself.

## Special Problems

The evolutionary study of molecules has raised a number of special problems, not always seen in the same way by molecular and organismal biologists. The phenomenon that has caused most trouble in attempts to determine evolutionary affinities is convergence: the development of similar characteristics by organisms of different ancestry. Any addition of evidence would be most welcome, especially if it involved characters unlikely to converge. Here the molecular biologists do not agree; Wald (*6*), for example, says that convergence is much more likely at the molecular level, while Zuckerkandl (*17*) independently maintains that it is less likely. To me, as an organismal biologist, it seems that Wald is probably right. Convergence to the point of identity or of seriously confusing similarity would appear to be more likely in a single kind of molecule, even one as complicated as a protein, than in such phenotypic characters as are end results of the interactions of a very large number of such molecules. Anfinsen (*7*) cited an example (from the work of Sanger *et al.*) indicating from insulin composition that sperm whales are identical with pigs and quite different from sei whales! (*19*). To be sure, a sequence of only three amino acids is involved, and both differences and resemblances could be incidental without even true convergence, but the lesson is there. Fortunately, the fact that protein and morphological convergence may be independent of each other gives a double check if the evidence of both is available.

Another problem, discussed at some length by Anfinsen (*7*), arises from the evidence that proteins have parts that can vary greatly or even be removed altogether without seeming to affect function. There is also the concept of "dormant genes" [discussed by Zuckerkandl (*17*), among others, and in studies which he cites; see also Zuckerkandl and Pauling (*20*)]. This concept is, again, related to the hypothesis of regular, secular change in molecules, mentioned in the opening paragraph of this article. Essentially the same question has long been discussed by evolutionary biologists, in this form: Can a gene (or allele) be neutral with respect to selection? (Much of the discussion is summarized, with citations, in *14*.) It is impossible to establish complete absence of exceptions, but so far every supposedly neutral gene that has been adequately investigated has turned out not to be neutral. There is a strong consensus that completely neutral genes or alleles must be very rare if they exist at all. To an evolutionary biologist it therefore seems highly improbable that proteins, supposedly fully determined by genes, should have nonfunctional parts, that dormant genes should exist over periods of generations, or that molecules should change in a regular but nonadaptive way.

This unsettled question could have far-reaching significance, for instance through the hypothesis [suggested but not fully supported by Anfinsen (*7*)] that the invariable or fully homologous parts of proteins in different animals are the functional, or at least the most significantly functional, parts. It would then seem to follow that the actual specific differences in proteins may be little or not at all adaptive, and this again seems unlikely to an organismal biologist. However, Anfinsen also points out (and the examples could be largely multiplied from other sources) that, for instance, serum proteins with no immunochemical similarity at all may be fully and identically functional. It is certainly not true as a generalization that molecular differences among species are commonly nonfunctional or nonadaptive, and indeed I think no molecular biologist would go to that extreme.

It is undoubtedly on questions related to adaptation that an evolutionary synthesis of molecular and organismal viewpoints and data will be most useful. I shall here give briefly two further examples from work by Wald (*6*, and earlier papers cited therein), not because I happen to disagree with his interpretations but because his brilliant studies provide such ideal data on the molecular basis of organismal adaptation. He shows that freshwater vertebrates generally have retinal pigments containing vitamin $A_2$, while marine and land vertebrates generally have $A_1$. He interprets this as a phylogenetic phenomenon, with $A_2$ in ancestral (true) fishes, supposedly freshwater forms, and $A_1$ developed in progressive phylogeny by marine and land descendants. He finds it inexplicable and almost an unnecessary complication that, for instance, reptiles, primitively having $A_1$, "revert" to $A_2$ when they adapt to fresh water. To an organismal biologist, the picture, including the apparent anomalies and supposed reversions, suggests interpretation in terms of adaptation, primarily, and phylogeny only secondarily. Many, but perhaps not quite all, of the observations would be explained if we assumed that $A_2$ is adaptive in freshwater forms and $A_1$, in land and saltwater forms—so much so that selection usually produced these adaptations rapidly and tended to erase purely phylogenetic effects. I have no idea what the difference in adaptation might be, but suggest that study from this point of view might clarify the molecular function involved.

A second example from Wald is his demonstration that tadpoles resemble fishes in a number of biochemical characteristics, whereas adult frogs have a biochemistry more like other land vertebrates. Amphibians were of course derived from fishes, and Wald interprets these changes as "the most striking instances we know of recapitulation." In my opinion there is no reason to invoke recapitulation and definite reason not to. As regards the species in question, it would appear that tadpoles are adapted to live in the water and adult frogs to live on land. In spite of some complications, this is the plausible explanation for nitrogen excretion: ammonia in water, urea out of it. Other changes may be less clearly adaptive but are likely, at least, to be adaptive. Some of the evidence, also given in part by Wald, is that when amphibians go from land to water, as some do, the changes tend to go in the opposite direction; they antirecapitulate!

## The Adaptive System

Finally, let us turn (or return) to the structure of the whole adaptive system, its causations, and the place of molecules in it. The most basic of all molecules, in this context at least, is DNA. Its influence is exerted, in part if not altogether, through RNA. Recognizing the RNA as an agent of DNA in this sequence, we conclude that RNA is not the cause of the eventual action: synthesis of a protein. (One could raise some delicate semantic problems here, but I think the statement can stand as written for present purposes.) Then is the DNA the causative agent in a really explanatory sense? It carries, as we say, a message (another semantic problem!) and is indeed a messenger and an agent just as much as messenger RNA is. In following the chain back we reach a really significant point of causation not when we locate the message, which is in the DNA, but when we learn where the message came from to begin with, what composed it. Any message composed, so to speak, by the DNA itself would be in the language of mutation. But mutations are predominantly inadaptive, and the message, beyond doubt, is almost entirely adaptive. Mutations form what may be called letters or words, to continue the now somewhat shopworn metaphor, and in that way they supply materials that permit something new to be said and that limit what can be said. However, they certainly do not compose the message in any meaningful sense.

The message, or at very least the greater part of it, relates to interaction of organism and environment. The interaction involves the whole organism, and hence arises and expands from the molecular level. There must be some sort of feedback from the organism-environment interaction into DNA, and hence into the other molecules. There are, as is well known, innumerable feedback mechanisms at the molecular level itself, and many or most of these are responsive to interactions with the environment. The Neo-Lamarckians, before much was known about feedback or anything at all was known about molecular genetics, supposed that evolutionary feedback was of the same kind, within individuals and into the genetic system, whatever that might prove to be. Now, however, we do know about DNA and other essentials of the genetic system, and we know beyond serious doubt, even though it seems rather odd, that DNA is not subject to feedback within individuals. That is, as Pontecorvo (*21*) has put it, "the *structure* of the genetic material is not subject to regulatory change . . . although the *expression* of the genetic material . . . is subject to regulation—qualitative and quantitative—at all levels of organization. . . ."

Changes in individual expression—to put it figuratively, the way the message is read—do not affect the message itself. The necessary message-constructing feedback is not here but in a system of higher order: in the population and not the individual. It operates through natural selection, which operates in populations, just as populations are what really evolve. Thus, through a different approach we come again to natural selection and now see it as the most truly causative (although not the only) element in the adaptive system. Viewed in this way, it is the composer of the genetic message, and DNA, RNA, enzymes, and the other molecules in the system are successively its messengers.

### References and Notes

1. M. Florkin, *Biochemical Evolution*, S. Morgulis, translator (Academic Press, New York, 1949).
2. C. A. Williams, Jr., and C. T. Wemmyss, Jr., "Experimental and evolutionary significance of similarities among serum protein antigens of man and lower primates," *Ann. N.Y. Acad Sci.* **94**, 77 (1961).
3. E. Margoliash, "Primary structure and evolution of cytochrome c," *Proc. Nat. Acad. Sci. U.S.* **50**, 672 (1963).
4. M. Goodman, "Man's place in the phylogeny of the primates as reflected in serum proteins," in *Classification and Human Evolution*, S. L. Washburn, Ed. (Wenner-Gren Foundation, New York, 1963).
5. R. L. Hill, J. Buettner-Janusch, V. Buettner-Janusch, "Evolution of hemoglobin in primates," *Proc. Nat. Acad. Sci.* **50**, 885 (1963).
6. O. Wald, "Phylogeny and ontogeny at the molecular level," in *Evolutionary Biochemistry*, A. I. Oparin, Ed. (Pergamon, London, 1963).
7. C. B. Anfinsen, *The Molecular Basis of Evolution* (Wiley, New York, 1959).
8. T. M. Sonneborn, "Implications of the new genetics for biology and man," *A.I.B.S. (Am. Inst. Biol. Sci.) Bull.* **1963**, 22 (Apr. 1963).
9. P. Weiss, "From cell to molecule," in *The Molecular Control of Cellular Activity*, J. M. Allen, Ed. (McGraw-Hill, New York, 1961).
10. P. T. Mora, "Urge and molecular biology," *Nature* **199**, 212 (1963).
11. C. S. Pittendrigh, "Adaptation, natural selection, and behavior," in *Behavior and Evolution*, A. Roe and G. G. Simpson, Eds. (Yale Univ. Press, New Haven, 1958).
12. V. Grant; *The Origin of Adaptations* (Columbia Univ. Press, New York, 1963).
13. G. G. Simpson, *This View of Life* (Harcourt, Brace & World, New York, 1964).
14. E. Mayr, *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, Mass., 1963).
15. H. F. Blum, "On the origin and evolution of living machines," *Amer. Scientist* **49**, 474 (1961).
16. E. Caspari, "Genetic basis of behavior," in *Behavior and Evolution*, A. Roe and G. G. Simpson, Eds. (Yale Univ. Press, New Haven, 1958).
17. E. Zuckerkandl, "Perspectives in molecular anthropology," in *Classification and Human Evolution*, S. L. Washburn, Ed. (Wenner-Gren Foundation, New York, 1963).
18. These systems have been discussed from different points of view by I. M. Lerner [*Genetic Homeostasis* (Oliver and Boyd, Edinburgh, 1954)]. I. I. Schmalhausen [*Factors of Evolution: The Theory of Stabilizing Selection* (Blakiston, Philadelphia, 1949)], C. H. Waddington [*The Strategy of the Genes* (Allen & Unwin, London, 1957)], and others.
19. As is all too common, the animals are not precisely specified. The pig is presumably the domesticated form of *Sus scrofa*. The sperm whale is *Physeter catodon*; the sei whale, *Balaenoptera borealis*. The two whales belong to different suborders of the same order. The pig belongs to a different order and cohort from the whales.
20. E. Zuckerkandl and L. Pauling, "Molecular disease, evolution and genetic heterogeneity," in *Horizons in Biochemistry*, M. Kasha and B. Pullman, Eds. (Academic Press, New York, 1962).
21. G. Pontecorvo, "Microbial genetics: retrospect and prospect," *Proc. Roy. Soc. London* **B158**, 1 (1963).
22. Drs. Ernst Mayr, George Wald, and Curtis Williams kindly read the manuscript and offered constructive suggestions; they do not necessarily endorse my opinions.