– 1 –

MACROMOLECULES AND NATURAL SELECTION

Francis H. C. Crick

UNIVERSITY OF CAMBRIDGE

There is a very real sense in which the nucleic acids and the proteins are the key molecules of living systems. This is not to deny that carbohydrates, lipids, coenzymes, and other small molecules are important. What criterion, then, justifies us in putting this emphasis on proteins and nucleic acids? If one leaves theory aside and notes simply what we observe, then the viruses provide the most telling evidence. Many viruses consist only of protein and nucleic acid and very little else; no natural virus exists without them. However, I wish to give theoretical reasons why this should be so, and this forces me to consider the basic properties of living systems.

Here again one cannot say that some properties are all-important while others are merely trivial. For example, a living system must obtain free energy from its environment; this necessarily involves metabolism, and it is not unreasonable to put metabolism as an essential property. But I wish to stress a different point of view. As I see it, a living system, as we find it today, exists only because it has evolved, and in the long run it will continue to "exist" (that is, to have descendants) only as long as it is capable of evolving further, or at least while it can counteract the unavoidable tendency to make "mistakes." In a word, a living system implies natural selection. It is thus worthwhile to consider what general properties are essential to a living system. We can then ask, in a naive way, whether the properties of the macromolecules concerned are related in any rather direct manner to these theoretical requirements. Surprisingly, it turns out that there does appear to be such a connection; it is this connection that I wish to examine.

Of course, it would be more prudent to defer such an examination until we were more certain of the roles of protein and nucleic acid—that is, until we had such detailed knowledge that we would run little risk of mistaking the principles involved. However, it seems to me that we can already make intelligent guesses about the sort of way the molecules function, and that it will do no harm to open the subject for discussion.

I shall assume, then, that there is some validity in the current ideas about molecular biology: that almost every biochemical process is catalyzed by a special enzyme; that all enzymes are proteins; that proteins (with the exception of cross-linking by S–S bonds) are unbranched polypeptides, made from a standard set of 20 amino acids; and that each one has its own precisely determined amino-acid sequence and is folded in a special manner which is essential for its activity. As to the nucleic acids, I shall follow the current "fashionable" ideas: that they constitute the most important part, if not the sole part, of the genetic material; that DNA is replicated by the complementary pairing mechanism; that the genetic information lies mainly in the exact sequence of bases of the nucleic acid; and that its main function is to control, in some way not yet understood, the amino-acid sequence of proteins. And I shall assume that the reader is tolerably familiar with all this (see the general references).

Let us first examine what properties we need for natural selection to operate. It is probably impossible to give an exhaustive set of abstract postulates; rather we will generalize from the type of system that nature has actually produced.

The first requirement appears to be for specific replication—that is, at some stage there must be a rather exact copying process. This need not necessarily be *direct: i.e., A* might produce a copy *B; B* a copy *C;* and then *C* a copy like *A*. But I do not think it can be entirely "arithmetical." By this I mean it is not enough just to have a master copy, from which are made subsidiary copies which cannot themselves be copied. It is not enough to have a printing press producing newspapers. There must be a mechanism by which the newspaper can be copied back into type. This kind of copying, in which a copy can itself be copied, I call "geometrical," and for natural selection to operate, the process must be "geometrical" to some extent. The reason is obvious. Natural selection has two functions. It is in part a device to enable errors to be corrected (especially errors of replication), and since some errors are inevitable, the population must be able to eliminate them. But this very process of protecting against a downward drift gives the necessary mechanism for a positive evolution—*i.e.*, not merely a stabilization of the status quo but a progressive improvement in "fitness." In order to do either of these jobs, the organism must replicate geometri-

cally so that errors can be eliminated without eventually diminishing the size of the population to zero.

The second requirement is that the mechanism should be able to copy a "mutation"—that is, to copy a special kind of mistake.

This property might not seem essential if an organism merely had to preserve the status quo, but it is necessary if the organism is to evolve. As I shall show, a rather simple way of doing this follows from the next requirement.

This third requirement is not quite as obvious as the first two, and yet it seems to me to be in the long run just as important. This is the requirement for "versatility." The population, if it is to survive in a hostile or indifferent world, must be able to perform a variety of functions, and in particular it must be able to carry out a great variety of chemical reactions. Thus the "genetic material"—the part of the organism that is copied geometrically—must be able to express itself chemically in many different ways; this is what I mean by "versatility." It obviously includes, among other things, the ability to metabolize.

It is not surprising, therefore, that in order to do this, organisms have evolved a "language." The genetic material must contain "information" to enable the organism to carry out its chemical acts, and it must contain a lot of it. A very efficient way of conveying information is to have a small number of different symbols, and to allow the linear *order* of the symbols to constitute the information: that is, by a language rather than, say, by a picture. The simplest copying process for a piece of such a language is one in which a copy is made letter by letter, without much reference to adjacent letters. This implies that if a mistake is made, and one letter is accidentally *substituted* for another, then this mistake is of such a type that it in its turn can be copied. This, as we saw, was the essential requirement for a "mutation" if it is to provide the raw material upon which selection can operate.

I think there is at least one further requirement for natural selection: it does seem as if the genetic material and at least some of its products must stay together in one place, so that they can act as a unit. This brings us to the idea of a "cell," with an outside and an inside; this implies membranes and, in molecular terms, lipids, but since they are outside my topic, I shall not pursue this further here.

Let us now look at the matter from the other angle—that of the macromolecules. Now I do not wish entirely to traverse familiar ground (for example that already covered in Crick, 1958). It is now widely appreciated that proteins and nucleic acids do indeed have some of the properties we expect of a language. Each is made from a small number of symbols (monomers) joined together in a linear order. The symbols (four for each of the nucleic acids, 20 for proteins) are universal throughout nature (with minor exceptions). For the proteins, at least,

the order of the monomers appears to be very precisely determined, and we have some grounds for believing that this is true for the nucleic acids. Thus if we concentrate on the *chemical* bonds of the macromolecules, we see that their "structure" is like that of a linear language.

What I want to consider here, however, is not the chemical bonding (their "primary structure") but the secondary and tertiary structure due to weaker bonds, such as hydrogen bonds, salt linkages, van der Waal forces, etc. In short, how they fold up and how the ways they can fold are related to their function.

Let us first consider the nucleic acids. We know rather little about the structure of RNA, but we have as evidence the double helix of DNA and the structure of the RNA-like polymers—the synthetic polyribotides—such as polyadenylic and polyuridylic acids, etc. An early review of the evidence is given in Crick (1957) and a more recent one by Rich (1959).

The remarkable fact emerges that so far there is no simple regular structure for nucleic acid *having only one chain.* We can have regular helices with two chains or with three chains wound helically around one another (so far none is known for certain to have four chains). Of course, a single chain might take up a "two-chain" type of structure by folding back on itself, but for the moment I am excluding such complications. Moreover, in all these structures, as far as we know, the chains are held together merely by weak bonds between bases on different chains.

When we come to consider proteins and synthetic polypeptides, we find just the opposite picture. No simple two-chain structure is known. We have one three-chain structure—that of collagen—but the chains are not held together very firmly, and there is some restriction on the amino-acid sequence (every third one must be glycine). In particular, interactions between the backbones are what mainly hold the structure together, though the interaction between pyrrolidine rings appears to help. It seems unlikely that the structure of collagen could ever be the basis for a simple, precise replication mechanism, based on the pairing of side-chains.

What we *do* find for polypeptides is that the simplest regular structure is of a *single* chain coiled helically on itself—the well-known alpha-helix—and we now have no doubt, thanks mainly to the work of Kendrew and his colleagues (1960), that the alpha-helix is important in globular proteins.

The significant point about the alpha-helix, however, appears to be its stability. The backbone itself is, in a loose sense, only marginally stable in water. That is, many sequences of side-chains will fold up into an alpha-helix, but it is possible, by the interaction of side-chains,

to interrupt the regular fold. This fact, and the existence of 20 types of side-chain rather than only four, allows the polypeptide chain to take up an enormous variety of different folds, depending at least in part on the amino-acid sequence. (Whether the fold depends on anything else remains to be seen.) Thus upon a rather simple chemical ground plan it is possible to build many different well-defined structures, and it is this that gives the proteins, as a class, their enormous versatility.

Can nucleic acid take up defined but complicated folds? Unfortunately we do not know, though we suspect not. What we do know is that the *functions* of nucleic acid appear rather limited (though this may, of course, merely reflect our ignorance), whereas the functions of proteins are extremely various and very delicately adjusted to any particular job. When we know just how DNA and RNA control the synthesis of proteins, as we believe they do, we shall be in a better position to assess the limitations imposed by the folding of polynucleotides.

It is easy to see, then, that with nucleic acid alone or with protein alone we would have great difficulty in meeting efficiently the demands of natural selection. With protein alone, the replication mechanism would have to be more complicated, since we cannot form a two-chain complementary structure—at least not, as far as we know, with the present side-chains. With nucleic acid alone, we could probably replicate quite nicely, but it would be difficult to use the nucleic acids as the basis for a vast family of enzymes. Life as we know it appears to be a symbiosis between these two very different families of polymers. Each —protein and nucleic acid—contributes its own particular capabilities, and the kinds of ways they can fold turn out to be a rather significant part of these capabilities.

When our knowledge of the biosynthesis and functions of these polymers is much more complete, and when we know the manner in which the nucleic acids and the proteins are linked symbiotically (the problems of protein synthesis, "coding," etc.), it is reasonable to expect that other key features of biology will be explainable by basic molecular properties. For example, I have said nothing about the necessity for "particulate" rather than "blending" inheritance. Again, it seems highly likely that the general absence of the inheritance of acquired characteristics can be explained by the *irreversibility* of RNA and protein synthesis, produced by the large flow of free energy into the process needed to minimize mistakes. At bottom, it is the simplicity and universality of the basic operations of biochemistry that encourage one to look for the explanation of the general features of biology in simple molecular properties.

References

For general references, the reader is referred to the articles in the following works:

1. Oncley, J. L., ed., "Biophysical Science—A Study Program," *Review of Modern Physics*, January and April, 1959. Also in book form as *Biophysical Science* (N. Y., Wiley, 1959).
2. *Structure and Function of Genetic Elements*, Brookhaven Symposium No. 12 (Upton, N. Y., Brookhaven National Laboratory, 1959).
3. Anfinson, C. B., *The Molecular Basis of Evolution* (N. Y., Wiley, 1959).

Crick, F. H. C., 1957. "The Structure of the Nucleic Acids and Related Substances," in Allfrey, V. G., *et al.*, eds., *Cellular Biology, Nucleic Acids and Viruses*, N. Y. Academy of Sciences Special Publication No. 5 (N. Y., N. Y. Acad. of Sci.).

Crick, F. H. C., 1958. "On Protein Synthesis," in "The Replication of Macromolecules," *Symp. Soc. Exp. Biol. 12*, 138.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C., 1960. "Structure of Myoglobin," *Nature 185*, 422.

Rich, A., 1959. "Polynucleotide Interactions and the Nucleic Acids," in *Structure and Function of Genetic Elements* (see reference 2, above), p. 17.