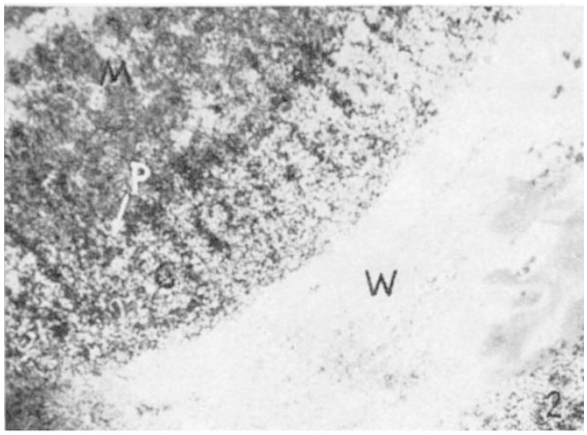


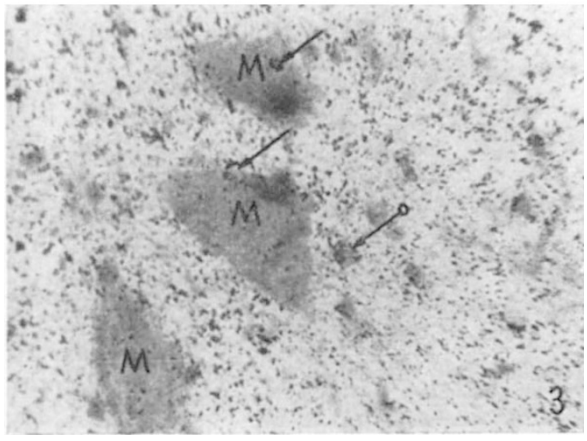


Fig. 2. Cerebellar cortex of the same rat as Fig. 1. Unstained autoradiogram of a cryostat section. M, Molecular layer; P, Purkinje cells (non-reacting); G, granular layer; W, white matter.

Fig. 3. Lumbar spinal cord (anterior horn) of a rat, 45 min after a single injection of TSC-^{14}C. Bouin-fixed and embedded specimen, counterstained with haematoxylin and eosin. Motoneurones (M) are non-reacting; capsular glial cells (arrow) and astrocytes in the grey matter (arrow with a circle) contain numerous grains.

binding by non-reacting structures might indicate those cellular elements that do not exert GAD activity. Such an approach is actually an adaptation of the principle described by Ostrovsky and Barnard[5].

Rats weighing 160 g were injected intraperitoneally with 5–20 mg/kg thiosemicarbazide-^{14}C (specific activity 10 mCi/mmole, obtained from the Central Isotope Institute, Budapest). Characteristic convulsions and lethal "jumps" appeared 45–120 min after injection. The animals were killed by decapitation and samples of the central nervous system (brain, medulla, cerebellum, spinal cord) as well as other tissues were either fixed on Bouin's solution and prepared for autoradiography in the usual way or were not fixed but frozen with dry ice, cut on a cryostat and applied to slides pre-coated with Kodak *AR*–10 stripping film, emulsion side up. The exposure time was 3–8 weeks; autoradiograms were developed in Kodak autoradiographic developer and some were counterstained with haematoxylin and eosin.

Fig. 1 shows a coronal section of the brain of a rat given two intraperitoneal injections of 10 mg/kg TSC-^{14}C, with an interval of 40 min between them, and then killed 90 min after the first injection. The heaviest reaction is confined to the hippocampus and fascia dentata chiefly in the layer of hippocampal pyramids (Fig. 1, inset); silver grains are, however, not found in the nerve cells themselves but are concentrated in the surrounding neuropil. Activity is high also in the nucleus habenularis medialis. There is a moderate reaction in the cortex, especially in the vicinity of the interhemispherical fissure.

Both the molecular layer and the granular layer of the cerebellum contain numerous silver grains (Fig. 2). The localization pattern of the reaction does not, however, conform to any of the neural elements but resembles more closely the glial structure. The reaction is slightly stronger around Purkinje cells, which are themselves devoid of any reaction both in pre-fixed and in non-fixed specimens. On the other hand, in deep cerebellar nuclei (and also in some of the brain stem nuclei, for example the substantia nigra and nuclei pontis) silver grains are located within the cytoplasms of nerve cells. Weak or virtually no activity can be seen in the white matter, although the glial cells in the white matter react if higher doses of TSC are used or if the animals are killed after a shorter time interval.

The nerve cells in the spinal cord do not contain silver grains, but capsular and other glial cells in both the grey and white matter contain numerous grains (Fig. 3). According to Curtis[2], GABA is not involved in spinal inhibitory mechanisms.

The gross distribution of silver grains reduced by TSC-^{14}C is in accord with the distribution of GAD and GABA anticipated on the basis of earlier biochemical and pharmacological studies (cerebellar cortex[2], hippocampus[2], substantia nigra[6]). While the concentration of silver grains in nerve cells (dentate nucleus, substantia nigra) and in the neuropil surrounding them (hippocampus) is consistent with current views, it is striking that Purkinje cells do not exert any reaction. Proteins sensitive to TSC are undoubtedly more widespread than GAD, so the apparent localization of TSC in glial cells of the cerebellar cortex may be partly due to binding of the drug to other B_6-dependent enzymes. The possibility cannot be excluded, however, that, at least in this area, GABA is produced by glial cells, perhaps in order to be taken up by nerve terminals and/or nerve cells in a second step.

Bertalan Csillik
Elizabeth Knyihár

Department of Anatomy,
University Medical School,
Szeged, Hungary.

Received August 12; revised October 6, 1969.

[1] Roberts, E., in *Structure and Function of Inhibitory Mechanisms*, 401 (Pergamon Press, Oxford, 1968).
[2] Curtis, D. R., in *Structure and Function of Inhibitory Mechanisms*, 429 (Pergamon Press, Oxford, 1968).
[3] Van Gelder, N. M., *J. Neurochem.*, **12**, 231, 239 (1965).
[4] Killam, K. F., Dasgupta, S. R., and Killam, E. K., in *Inhibition in the Nervous System and Gamma-Aminobutyric Acid*, 302 (Pergamon Press, Oxford, 1960).
[5] Ostrovsky, K., and Barnard, E. A., *Exp. Cell Res.*, **25**, 456 (1961).
[6] Albers, R. W., in *Inhibition in the Nervous System and Gamma-Aminobutyric Acid*, 196 (Pergamon Press, Oxford, 1960).

Natural Selection and the Concept of a Protein Space

Salisbury[1] has argued that there is an apparent contradiction between two fundamental concepts of biology—the belief that the gene is a unique sequence of nucleotides whose function it is to determine the sequence of amino-acids in a protein, and the theory of evolution by natural selection. In brief, he calculated that the number of possible amino-acid sequences is greater by many orders of magnitude than the number of proteins which could have existed on Earth since the origin of life, and hence that functionally effective proteins have a vanishingly small chance of arising by mutation. Natural selection is

© 1970 Nature Publishing Group

therefore ineffective because it lacks the essential raw material—favourable mutations.

I should like to look at the problem from a different point of view. I shall assume that mutations, while not random in a chemical sense, are random as far as their chances of improving the function of the corresponding proteins are concerned. I shall also assume that evolution has occurred either by the natural selection of favourable mutations or by the chance fixation by genetic drift of selectively neutral mutations. The justification for making these assumptions is that no sensible alternatives have been suggested and that no evidence exists at the moment to invalidate them. If these assumptions are true, what can we say about the frequency and distribution of amino-acid sequences which are functional, either as enzymes or in some other way?

The model of protein evolution I want to discuss is best understood by analogy with a popular word game. The object of the game is to pass from one word to another of the same length by changing one letter at a time, with the requirement that all the intermediate words are meaningful in the same language. Thus WORD can be converted into GENE in the minimum number of steps, as follows:

WORD WORE GORE GONE GENE

This is an analogue of evolution, in which the words represent proteins; the letters represent amino-acids; the alteration of a single letter corresponds to the simplest evolutionary step, the substitution of one amino-acid for another; and the requirement of meaning corresponds to the requirement that each unit step in evolution should be from one functional protein to another. The reason for the last requirement is as follows: suppose that a protein A B C D . . . exists, and that a protein a b C D . . . would be favoured by selection if it arose. Suppose further that the intermediates a B C D . . . and A b C D . . . are non-functional. These forms would arise by mutation, but would usually be eliminated by selection before a second mutation could occur. The double step from a b C D . . . to A B C D would thus be very unlikely to occur. Such double steps with unfavourable intermediates may occasionally occur, but are probably too rare to be important in evolution.

This is a model of the way in which one gene may change into another. An increase in the number of different genes in a single organism presumably occurs by the duplication of an already existing gene followed by divergence. If so, it remains true that new genes arise as modifications of pre-existing ones.

It follows that if evolution by natural selection is to occur, functional proteins must form a continuous network which can be traversed by unit mutational steps without passing through nonfunctional intermediates. In this respect, functional proteins resemble four-letter words in the English language, rather than eight-letter words, for the latter form a series of small isolated islands in a sea of nonsense sequences. Of course, this is not to deny the existence of isolated island proteins, analogous to the four-letter words ALSO and ALTO.

It is easy to state the condition which must be satisfied if meaningful proteins are to form a network. Let X be a meaningful protein. Let N be the number of proteins which can be derived from X by a unit mutational step, and f the fraction of these which are meaningful, in the sense of being as good as or better than X in some environment. Then, if $fN > 1$, meaningful proteins will form a network, and evolution by natural selection is possible. In estimating N it is necessary to distinguish two classes of mutations: (i) substitutions of single amino-acids, and additions or deletions of small numbers of amino-acids, making only a small change to the protein; and (ii) mutations producing a major change in amino-acid sequence, such as frame shifts and intramolecular inversions.

Mutations of the former type are much more likely to give rise to meaningful proteins than the latter. In the same way, a single random letter substitution in a meaningful word is more likely to give rise to a meaningful word than the simultaneous alteration of all the letters. Although frame shift mutations are known to occur, it is not clear whether they have ever been incorporated in evolution. It is therefore better to take N as the number of possible substitutions of single amino-acids. If all substitutions were possible in a single mutational step, N for a protein of 100 amino-acids would be 1,900. In practice the genetic code limits N to approximately 10^3.

Hence f must be greater than 1/1,000. It does not follow that the fraction of all possible sequences which are meaningful need be as high as 1/1,000. It is probably much lower. There is almost certainly a higher probability that a sequence will be meaningful if it is a neighbour of an existing functional protein than if it is selected at random. In fact, in treating N as the number of amino-acid substitutions rather than as the total number of possible mutational steps, it was in effect assumed that a random sequence has a negligibly small probability of being functional; this assumption will be confirmed if it turns out that frame shifts are rarely or never incorporated in evolution.

Suppose now that we imagine all possible amino-acid sequences to be arranged in a "protein space", so that two sequences are neighbours if one can be converted into another by a single amino-acid substitution. Then the requirement that $fN > 1$ requires that the "density" of functional proteins in certain regions of the space must be quite high—perhaps greater than 1/1,000. This agrees with Salisbury's conclusion that proteins, and hence the genes that determine them, cannot be as unique as all that. As a convinced Darwinist, I published[2] the conclusion that $fN > 1$ when little was known about the frequency of amino-acid substitutions in evolution. Since then evidence has accumulated (for a review, see King and Jukes[3]) that many substitutions are either selectively neutral or at least make comparatively minor changes in the function of proteins.

If $fN > 1$, no quantitative difficulty arises in explaining the evolution of proteins by natural selection. A difficulty nevertheless remains in explaining the origin of life—that is, in explaining the origin of the first functional proteins together with the genetic mechanism for producing them. If it were true that only a minute fraction of possible amino-acid sequences have even the slightest enzymatic activity, it would be difficult to understand how the first proteins arose. I do not want to discuss the problem of the origin of life, but only to point out that it is a quite different problem from that of the mechanism of evolution.

Some questions about molecular evolution can be formulated more clearly in terms of a protein space. For example: (i) Are all existing proteins part of the same continuous network, and if so, have they all been reached from a single starting point? Possible alternatives are that there are two or more distinct networks, or that there is one network with multiple starting points. (ii) How often, if ever, has evolution passed through a non-functional sequence? If so, has this been achieved by the random walk of genes rendered redundant by duplication, or by the chance concurrence of two or more mutations? (iii) What fraction of the functional network has already been explored in evolution? (iv) What fraction of potentially useful proteins are inaccessible?

JOHN MAYNARD SMITH

School of Biological Sciences,
University of Sussex.

Received November 7, 1969.

[1] Salisbury, F. B., *Nature*, **224**, 342 (1969).
[2] Maynard Smith, J., in *The Scientist Speculates* (edit. by Good, I. J.) (Heinemann, London, 1961).
[3] King, J. L., and Jukes, T. H., *Science*, **164**, 788 (1969).

© 1970 Nature Publishing Group