made the observations, the relative intensity of the absorptions in spectra of the dark surface markings was about two, and of the light markings (which are thought to be deserts), less than 0·5. It seems unlikely that the absorptions detected in the spectra of the desert regions are caused by vegetation. But terpenes in the atmosphere above the deserts would give just such weak absorptions.

Hence the terpene-haze hypothesis of the Martian atmosphere leads to a prediction that is capable of experimental check. If an infra-red spectrum of the Martian deserts can be obtained during one of the rare 'blue clearances', there should be no absorptions at 3·5 microns.

REFERENCES

BRIGGS, M. H.: *Spaceflight*, volume 2, page 237, 1960.
BRIGGS, M. H. and REVILL, J. P.: *J. Brit. Interplanetary Soc.*, volume 17, page 391, 1960.
RANKAMA, K.: *Geol. Soc. America, Special Paper*, volume 62, page 651, 1955.
HESS, S. L.: *Trans. N.Y. Acad. Sci.*, volume 19, page 352, 1957.
SINTON, W. M.: *Science*, volume 130, page 1234, 1959.
UREY, H. C.: *The Planets: Their Origin and Development*: New Haven, 1952.
WENT, F. W.: *Proc. Nat. Acad. Sci.*, U.S., volume 46, page 212, 1960.

87   The Limitations of Molecular Evolution
by JOHN MAYNARD SMITH

The analogy between messages in words and genetic instructions on chromosomes is not new, but there is one application of it which has I think been missed, and which has some amusing consequences. This is the analogy between molecular evolution and a particular word-game. In this game, it is required

to convert one word into another by changing one letter at a time, with the limitation that each intermediate stage shall also be a meaningful word. Thus WORD can be changed to GENE as follows:

W O R D
W O R E
G O R E
G O N E
G E N E

In evolution the unit events, or mutations, are changes in the sequence of the four bases in DNA, and are believed to have as their immediate consequence changes in the sequence of amino-acids in the proteins programmed by that DNA. At the protein level, probably the commonest evolutionary event is the substitution of one amino-acid for another, corresponding to a single event in the word game; other possible unit events are discussed below.

I suggest that a population of organisms which cannot now produce a particular protein X cannot evolve the capacity to do so unless X is connected by a series of unit steps (consequent on unit changes in DNA) to a protein it can now make, such that each intermediate is a functional protein which would, in some environment, be an improvement on, or at least as good as, its predecessor in the series. This is because any particular step will occur with very low frequency, so that the simultaneous occurrence of several favourable steps affecting the same protein molecule is ruled out. Hence each protein arising by a single step must increase in frequency in the population under the influence of natural selection before the next step can be taken.

Consider the evolution of proteins 50 amino-acids long. Since there are 20 different kinds of amino-acid commonly present in proteins, there are 20^{50} different possible sequences. What proportion of them would be biologically useful proteins I don't

know, but it seems likely that most of them would be useless.[1] If so, the functional proteins may occur as a series of 'islands' surrounded by a 'sea' of meaningless sequences, just as the words ALLY, ABLY, ABLE, AXLE form an isolated island in the word game. [ABBE, ABYE, and ALLA need not count. *G.Ed.*]

If f is the fraction of amino-acid sequences which are meaningful, in the sense of being functional proteins which would be favoured by natural selection in at least some environment, and if N is the number of transformations of a protein which can occur by unit mutational steps, then if fN is greater than unity the 'landscape' will be a continuous network of meaningful land together with a number of small isolated islands; but if fN is less than unity, the landscape will be a sea with many small islands.

The number N depends on the as yet unknown nature of the genetic code whereby base sequences in DNA determine amino-acid sequences in proteins. So far as we know, the unit changes possible in base sequence in DNA are:

(i)   the substitution of one base for another; GORE – MORE

(ii) the elimination of one or more consecutive bases; GORE – ORE.

(iii) the duplication of one or more consecutive bases; GORE – GOGORE.

(iv) the inversion of a group of bases; GORE – OGRE.

Types (ii) and (iii) would cause changes in the number of amino-acids in the protein, and could be involved in the evolution of one protein to another of the same length only by rather roundabout routes.

Type (i) can lead to the substitution of one amino-acid for another. Since there are 20 amino-acids, it might seem that there are 50 x 19 possible transformations of a sequence 50 units long. But this depends on the nature of the code. For example, according to the code suggested by Crick *et al.*[2] each amino-acid is determined by a sequence of three bases. Of the 64 possible

—

triplets, only 20 make sense (i.e. determine an amino-acid), the rest are nonsense. A meaningful triplet can be transformed into an average of 5 others by a single base substitution; thus ACC can change to BCC, ADC, ACA, ACB, but all other changes give nonsense triplets. This would give $N = 250$ approx. for changes of type (i). There will be some additional transformations of type (iv), involving one, or several neighbouring amino-acids (for example, ACA BCB and ACB ACB are both pairs of sense triplets in Crick's code); but long inversions will always be nonsense, because some meaningful triplets are transformed by inversion into nonsense triplets. I have not enumerated the possible inversion transformations for Crick's code, but it seems possible that for this code, allowing for transformations of types (i) and (iv), N is about 500.

Since we know that evolution has in fact taken place, there would seem to be two possibilities. Either there are types of mutational event in addition to those discussed here (for example, the presence of additional DNA with an unusually high mutation rate, not programming proteins necessary for survival, but throwing up a variety of sequences on the off-chance that one might be useful),[3] or fN is greater than unity. But even if the latter is true, there are likely to be isolated islands, comparable to the AXLE group in the word game, consisting of proteins which cannot be produced by evolutionary processes even though they would be selectively advantageous if they arose.[4] These proteins are the analogue of the wheel and the magnet in structural evolution. If they exist, perhaps we shall learn to synthesise the necessary DNA sequences and incorporate them in our domestic animals, to enable them to browse on and digest the nylon and polythene which we may expect by then to cover England to a depth of several feet.

REFERENCES AND NOTES

1. The example of sickle-cell anaemia suggests that the proportion

of proteins which are useful may be much higher in diploids – one good reason for being diploid.

2. CRICK, F. H. C., GRIFFITH, J. S. and ORGEL, L. E.: 'Codes without commas'. *Proc. Nat. Acad. Sci.*, volume 43, page 416, 1957. While this book was in the press, an article appeared (CRICK, F. H. C., BARNETT, L., BRENNER, S., and WATTS-TOBIN, R. J., *Nature*, 30 December, 1961) indicating a different 'triple' code which implies perhaps twice as large a value for N.

3. Large duplications occur, and provide stretches of DNA whose programme is not necessary for survival, because the message is already carried elsewhere. Such duplications can be repro-grammed by selection to determine new proteins. This is certainly an important evolutionary process, but it does not greatly help in crossing the 'seas' of meaningless sequences, because the initial duplicated sequence will be a meaningful sequence; if it were not, it would not have been present to be duplicated. The reason for suggesting DNA with an unusually high mutation rate is that in such DNA several favourable steps might be taken simultaneously.

4. If fN exceeds unity, then the equation $p = \exp[(p-1)fN]$ has a solution, less than unity, giving the proportion of viable proteins that are inaccessible. *G. Ed.*

88   Random Synthesis and Subsequent Separation

by N. W. PIRIE

There are two ways of getting a prearranged sequence of objects, for example playing-cards. They can be shuffled and dealt at random over and over again until the right sequence emerges; they can be looked at and arranged deliberately. Which course is adopted depends on the state of development of techniques for dealing and recognising the right sequence, or for arranging. Hitherto, chemical synthesis of large molecules has set out deliberately to put them together step by step in the right order, but techniques for separating complex mixtures are now reaching a state of development that makes the alternative approach possible. It may well be that this is how specific