J. theor. Biol. (1976) **60**, 413–419

Interactive Macromolecular Sites
II. Role in Prebiotic Macromolecular Selection and Early Cellular Evolution

NORMAN G. ANDERSON

South Carolina Memorial Cancer Institute, and Department of Surgery, The Medical University of South Carolina, Charleston, S.C. 29401, *U.S.A.*

(*Received* 15 *April* 1975, *and in revised form* 6 *October* 1975)

All cellular macromolecules are covered with sites or reactive domains to which binding sites can be made on other macromolecules. It is proposed that molecular selection began during prebiotic evolution when polymers which adhered together in three dimensional arrays sedimented out of solution. Those tending to interact two dimensionally to form membranes would be much less actively removed, those forming linear structures less so, and the soluble molecules remaining would tend to be those whose interactive counterparts had been removed. Concentration of the suspension thus produced would tend to yield vesicles having surface and possibly internal membranes, having some internal linear aggregates, and containing soluble macromolecules. Subsequent evolution would select against new surfaces which interacted strongly with those already present (unless an advantageous structure is formed), and would, through random variation, gradually yield a large collection of surfaces whose counter sites are absent. The limiting number of such sites is termed a perfect set and contains all of the information required to recognize foreignness. It is therefore a "self" at the molecular level.

1. Introduction

Beginning with the classical treatise of Oparin (1938), rational explanations of the prebiotic origin of both low (Miller, 1953; Ponnamperuma & Mack, 1965; Oro, 1961; Harada & Fox, 1964; Bar-Nun, Bar-Nun, Bauer & Sagan, 1970; Dose, Fox, Deborin & Pavlovskaya, 1974) and high (Dose *et al.*, 1974; Fox, 1964; Fox, Harada & Kendrick, 1959; Fox, 1965; Kimball & Oro, 1971; Rohlfing & Oparin, 1972) molecular weight organic compounds have been presented. The problem of how large molecules might be concentrated to form protocells has been repeatedly examined and it has been proposed that coacervate formation (Oparin, 1938; Oparin, 1971), adsorption

onto inorganic surfaces (Bernal, 1951) or formation of compact proteinoid microspheres (Fox *et al.*, 1959; Fox & Waehneldt, 1968) or evaporation were causally involved. It has been explicitly assumed that selective mechanisms in the sense that the term is used in cellular evolution were not operative until after the formation of the first "living" cells or self-replicating molecules (Orgel, 1973). In this paper we examine certain extensions of the theory of perfect sets (Anderson, 1976) and propose that paratactic selection antedated the formation of self-duplicating cells, and that the selection mechanisms proposed yielded the molecular species most likely to form such cells.

The major steps proposed here in the transition to life are (a) prebiotic synthesis of monomers, (b) formation of high molecular weight polymers from monomers, (c) fractionation with selection of most-likely-candidate macromolecules, (d) concentration of these followed by (e) encapsulation and (f) the appearance of autoreplication. Fractionation (molecular selection) and concentration may or may not be part of the same process.

The strategy is to break an extremely complex and, at first sight, apparently very improbable event—the evolution of living systems from primeval matter—into a long series of steps each of which initially appears reasonable and in many cases, on closer examination, inevitable. This has led, not only to useful insights, but to interesting predictions and experiments.

The first step, the prebiotic synthesis of monomers, continues to be explored both theoretically and experimentally and a variety of synthetic mechanisms are thought to have been possible. Under the conditions believed to have existed in the early atmosphere and seas, synthesis of both carboxylic acids and amines would have occurred. However, a key feature was alkalinity because in an alkaline environment carboxylic acids would be trapped in the liquid phase at once, while low molecular weight amines would tend to be volatile and would be recycled until rendered non-volatile by carboxylation to amino acids. This sets the stage for the development of organisms composed largely of amino acids, derivatives of non-volatile purine and pyrimidine bases and sugars, and using as metabolic intermediates organic acids (Anderson, 1963). To such organisms organic amines are generally quite toxic in any but trace amounts (Guggenheim, 1951) although they may have important functions at trace levels. Thus the pattern of contemporary biochemistry is seen as an inevitable result of early conditions. (It is interesting to speculate on prebiotic evolution in acid seas which could have given rise to an intermediary metabolism based on amines, with carboxylic acids being toxic.)

Protein synthesis presents more difficult problems, both in theory and in practice because synthetic long-chain peptides are almost always insoluble and form precipitating aggregates in aqueous solution.

2. Mechanisms of Aggregation

Two basically different mechanisms (with possible intergrades) of protein–protein interaction or aggregation are known. The first is that occurring when proteins are denatured so that hydrophobic groups are externalized. Such hydrophobic groups tend to render the proteins insoluble in aqueous solutions. These reactions appear to be truly non-specific in the sense that any two hydrophobic surfaces prefer each other to water. Most hydrophobic residues in native proteins are internal and usually contribute to the stabilization of tertiary structure. Unfolding or denaturation to externalize them will often result in irreversible aggregation. It appears that in many proteins the denatured aggregated state is not necessarily more stable, rather that interaction between hydrophobic areas stabilizes the proteins in an unfolded state. This appears to account for the stability of enzymes, antigens, and antibodies when covalently attached to solid supports. Individual bound proteins are too far apart to interact when unfolded and hence can return unhindered to their native configuration (Anderson *et al.*, 1975). The second mechanism of protein–protein interaction involves stereo-specific interactive sites on soluble proteins. These involve the summation of weak secondary forces and the removal of water from between sites or domains which physically match. Such interactions are termed specific, and it is unlikely that any non-specific interactions between large soluble molecules exist, as has been recently suggested by Boyd (1974).

3. Prebiotic Macromolecular Selection

During the synthesis of early macromolecules by whatever means, those aggregating and interacting through hydrophobic interactions to form three-dimensional aggregates precipitated and were sedimented to the bottom of the seas or ponds in which they were formed. This constitutes the first and a continuing prebiotic selection mechanism. It tended to leave in solution molecules which were more soluble in the ordinary sense.

Many of these also interacted, however, for reasons previously presented and reviewed briefly here. Soluble proteins are covered with highly interactive residues, which can interact through salt bridges, hydrogen bonds, or other weak forces. These residues may be grouped to form interactive sites to which specific countersites can be made. The size of interactive sites is limited by the radius of curvature of the proteins involved, and a site is only considered to be interactive if the countersite can exist on another protein. Sites to which no counter attaching site can be made are considered

to be blank sites. Interactive sites therefore may be considered to exist (potentially) in pairs. It is evident that some sites, on a purely statistical basis, have more chance of occurring on a prebiotically synthesized soluble protein than others. Some proteins, having many reactive sites, will tend to be part of large sedimenting particles, and so will disappear from the solution. Note that if each member of a reactive pair occurs, but in different amounts, and if interaction causes precipitation, then all of the member present in smallest amount, and part of the member present in larger amount will tend to be precipitated, resulting in the selection of one site as fit to remain in solution. As discussed previously, a set of proteins exhibiting only one member of each pair of the universe of interactive pairs is termed a perfect set. During prebiotic evolution there will, therefore, be a tendency to evolve solutions containing perfect sets, that is, solutions of proteins having minimal interactive tendencies when in dilute solution. The greatest selective removal will be for proteins forming three-dimensional aggregates or solid precipitates. Those forming two-dimensional aggregates or membranes will be very much less actively removed, linear aggregates still less so, and there will be no selective removal of those proteins which find no paratactic or interactive counterpart in the solution at all.

The result is a collection of molecules which, when concentrated further possibly by evaporation, could automatically tend to form small membrane-bound spherules containing soluble proteins.

The role of nucleic acids in the earliest protocells is still being debated. However, it is interesting to recall very early studies (Greenstein & Hoyer, 1950) demonstrating that nucleic acids are quite effective in preventing the heat denaturation of proteins. One role for them could well have been to encourage "renaturation" of thermally synthesized protein.

These considerations assist in getting past the difficult problem of the selection of advantageous proteins to incorporate into protocells, but unfortunately say nothing about how self-duplication may have arisen. We therefore skip from prebiotic selection and formation of membrane-bound spherules to the first self-duplicating systems, and ask how the theory of perfect sets may be applied to subsequent stages of early protein evolution.

4. Rules Governing Early Protein Evolution

Given a self-duplicating system containing one or more proteins, what rules govern the appearance of each new protein?

First, the new protein must have a function of its own which contributes positively to cell survival.

Secondly, the protein must not interact with any active sites on any protein already in the cell. It must not be an inhibitor of some important intracellular enzyme for example. It must also not react to diminish its own importance. For example, an enzyme could have an active site on one portion of its surface, and an auto-inhibitory site somewhere else on the same molecule, thereby inhibiting other members of its own species.

Each time a new potentially interactive site is adopted, its opposite member is forbidden unless that opposite member is important to structure formation. If N pairs of sites exist in the universe of all interactive sites, then a perfect set would contain N sites, or one member of each pair. If in addition both members of some pairs are present, and the number of such pairs is S, then the cell will contain $N+S$ interactive sites.

Early in evolution, and probably in all living prokaryotes the limit $N+S$ was not reached. However, the restrictions described would still prevail. When few surfaces are present the restriction is of little importance; only as completeness begins to be approached is variability markedly restricted.

Thirdly, as the total number of potentially interactive sites in a cell enlarges not only are there restrictions on new sites, but mutational options for existing proteins diminish markedly.

For plant and bacterial cells with cell walls or strong surface coatings little advantage is gained by approaching the perfect set limits. This is not true for the first animal cells which began to feed on other living cells because of the necessity to distinguish food organisms from self. An amoeba is not well served if its food-seeking activities resulted in the ingestion of its own pseudopod or a daughter cell after division. Rather some general simple mechanism for recognizing foreign cells or bits of inanimate food is required and is automatically supplied by approaching the complexity of a perfect set.

5. Discussion

Proteins, and other cellular macromolecules, possess large numbers of surfaces to which matching counter or binding sites can be made. The whole of cell organization rests on this principle and both cell structure formation and metabolic activity depend on specific molecular surfaces. Evidence for this specificity derives from a variety of sources and is reviewed in detail in a subsequent paper. For the purposes of this discussion it is considered that the number of such pairs of interactive sites is limited, and that adoption of one member of a pair by a cell precludes adoption of the other unless the pair serves a useful structure-forming function. This restriction applies both

to the evolving cell and to the antecedent macromolecular suspension from which cells first evolved.

It is further proposed that during prebiotic evolution polypeptides having three-dimensional binding capabilities similar to those of denatured proteins aggregated and were removed from solution. Those which tended to form two-dimensional structures such as membranes would be less actively removed, while linear aggregates and freely soluble proteins would tend to remain in solution. It must be stressed that living systems are equilibrium systems and no functional solid accretion structures exist in them. (Storage granules are not considered to be functional.) Rather all proteins appear to be in equilibrium with their counterparts in solution and can turn over at some rate even though the rate may be a very slow one. Hence some protein surfaces are chosen for their specific interactive surfaces, while, as postulated here, the remainder are chosen to be non-interactive. Given a limited universe of possible interactive pairs, there must exist a limit to the number of half pairs adopted in a cell during evolution and that limit has been termed a perfect set.

How real are these considerations? First, there can be no doubt that a large number of surfaces are forbidden in any given cell. The phenomenon of immunity and the existence of many highly toxic proteins support this view. For example, if the interactive site on abrin, ricin, or diphtheria toxin appeared on a mutant human cell protein it is quite doubtful that the cell would survive. Paratactic exclusion must be accepted as a reality.

The problem of how far cells are able to go in the direction of a perfect set is a more difficult one and would require long periods of evolution to approach. However, for animal cells to evolve which could feed on a large variety of other cells and cell products, and recognize them as food, some rather large fraction of a perfect set would be required. This assumes that the sharper a molecular self is defined—and it has been defined in these studies as a perfect set—the wider is the selection of surfaces reacted with and "recognized" as foreign.

As discussed in the previous paper, the range of ability to recognize foreignness and the allowable variability in cell protein structure are reciprocally related. As a perfect set is approached, a wider range of substances may be reacted with, ultimately approaching all non-self proteins. This at the same time very severely restricts variability of proteins native to the cell, because any mutation may alter a site so that it reacts with one already existing in the cell.

Experimentally, it is difficult if not impossible to discover exactly what fraction of a perfect set is possessed by one cell. However, the fact that amino acid substitutions thought to be neutral with respect to the function

of a particular enzyme occur at a very slow rate, and that very few alleles are seen for proteins which appear to be equally efficient when many different substitutions are made argues strongly for the existence of perfect sets. If they exist, then almost any alteration in the configuration of one protein would require a corresponding alteration in another to prevent a new interaction. Such double simultaneous alterations would occur at a very slow rate.

As will be discussed in a subsequent paper, the evolution of complex multicellular organisms may require a genome not limited to one perfect set, but rather containing information sufficient for many slightly different ones, each found in a separate differentiated cell.

REFERENCES

ANDERSON, N. G. (1963). In *Lectures in Biological Sciences* (J. I. Townsend, ed.), pp. 1–13. Knoxville: The University of Tennessee Press.

ANDERSON, N. G. (1976). *J. theor. Biol.* **60,** 401.

ANDERSON, N. G., WILLIS, D. D., HOLLADAY, D. W., CATON, J. E., HOLLEMAN, J. W., EVELEIGH, J. W., ATTRILL, J. E., BALL, F. L. & ANDERSON, N. L. (1975). *Anal. Biochem.*, **68,** 371.

BAR-NUN, A., BAR-NUN, N., BAUER, S. H. & SAGAN, C. (1970). *Science, N.Y.* **168,** 470.

BERNAL, J. D. (1951). *The Physical Basis of Life.* London: Routledge, Kegan Paul.

BOYD, W. C. (1974). *Experientia* **30,** 1473.

DOSE, K., FOX, S. W., DEBORIN, G. A. & PAVLOVSKAYA, T. E. (eds) (1974). *The Origin of Life and Evolutionary Biochemistry.* New York: Plenum Press.

FOX, S. W., HARADA, K. & KENDRICK, J. (1959). *Science, N.Y.* **129,** 1221.

FOX, S. W. (1964). *Nature, Lond.* **201,** 336.

FOX, S. W. (ed.) (1965). *The Origins of Prebiological Systems.* New York: Academic Press.

FOX, S. W. & WAEHNELDT, T. V. (1968). *Biochem. biophys. Acta* **160,** 246.

GREENSTEIN, J. P. & HOYER, M. L. (1950). *J. biol. Chem.* **182,** 457.

GUGGENHEIM, M. (1951). *Die biogenen Amine.* New York: S. Karger.

HARADA, K. & FOX, S. W. (1964). *Nature, Lond.* **201,** 335.

KIMBALL, A. P. & ORO, J. (eds) (1971). *Prebiotic and Biochemical Evolution.* Amsterdam, London and New York: North-Holland/American Elsevier.

MILLER, S. L. (1953). *Science, N.Y.* **117,** 528.

OPARIN, A. I. (1938). *The Origin of Life.* Translated by S. Margulis. New York: Dover Publications. 1965 edn.

OPARIN, A. I. (1971). *Scientia* **1,** 12.

ORGEL, L. E. (1973). *The Origins of Life,* p. 145. New York: John Wiley & Sons.

ORO, J. (1961). *Nature, Lond.* **191,** 1193.

PONNAMPERUMA, C. & MACK, R. (1965). *Science, N.Y.* **148,** 1221.

ROHLFING, D. L. & OPARIN, A. I. (eds.) (1972). *Molecular Evolution.* New York: Plenum Press.