Journal of Chromatography, 215 (1981) 115–120
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

CHROM. 13,816

COPOLYAMINO ACID FRACTIONATION AND PROTOBIOCHEMISTRY

SIDNEY W. FOX
Institute for Molecular and Cellular Evolution, University of Miami, Coral Gables, FL 33134 (U.S.A.)

SUMMARY

Investigation of the origins of living things by the uniquely appropriate method of successive approximation in attempted retracement of steps in molecular evolution has yielded: a comprehensive theoretical flowsheet from archaic inanimate matter to an infrastructured, microscopic, protoreproductive, protometabolic protocell; a laboratory model of the same; and an explanatory assessment of the natural variation component of Darwinian evolution. For each of these, the significance is dependent upon awareness of the intrinsic tendency of amino acids, in mixed sets, to order themselves. Without such awareness, it is believed these vistas would have been delayed for decades. Selfordering would have in turn been difficult to recognize and support were it not for the chromatographic developments in fractionation of copolyamino acids.

INTRODUCTION

Modern biochemical investigation would be unthinkable without the methods of fractionation that are indigenous to the craft. Especially useful among these have been procedures of the type pioneered by Porath[1]. The worth of chromatographic methods is well known to biochemists; however, the contributions of such methods to protobiochemistry, *i.e.* the theory of the origin of life, is not so widely perceived. The value of fractionation methods to the problem area of protobiochemistry is indirect. These methods first described a constrained heterogeneity in simulated precellular proteins. This kind of result in turn suggested that those proteins could have arisen before nucleic acids were available within primitive cells.

The heart of the problem of protobiochemistry has been the answer to the chicken–egg problem of biochemistry: which originated first, nucleic acids or proteins? This problem was expressed by Popper[2] as:.

"What makes the origin of life and of the genetic code a disturbing riddle is this: the genetic code is without any biological function unless it is translated; that is, unless it leads to the synthesis of the proteins whose structure is laid down by the code. But, as Monod points out, the machinery by which the cell [at least the non-primitive cell which is the only one we know] translates the code 'consists of at least fifty macromolecular components *which are themselves coded in DNA* (Monod[3]). Thus the code cannot be translated except by using certain products of its translation.

0021-9673/81/0000–0000/$02.50

This constitutes a really baffling circle: a vicious circle, it seems, for any attempt to form a model, or theory, of the genesis of the genetic code."

However, an essential distinction must be made between the modern requirements and the primordial parameters. In fact, in seeking to correct misunderstandings of the central dogma in molecular biology[4], Crick has stated that this tenet "was intended to apply only to present-day organisms, and not to events in the remote past, such as the origin of life or the origin of the code."

Without evidently having before him any data on the nature of primordial proteins, Monod asserted that the basic primary structure of the first functional protein[5] "discloses nothing other than the pure randomness of its origin." Since Monod recognized no relevant data, the randomness referred to in his statement was an assumption[6].

One potential answer to the chicken–egg question has been the postulate that nucleic acids might have, when first formed, catalyzed their own replication. This they do not do now, and no one has ever demonstrated such catalytic power in nucleic acids, although it has been sought. Such phrases as the *self*-replication of nucleic acid are, as Dillon[7] states, misleading.

On the other hand, a comprehensive substitute for the function of nucleic acids in blueprinting ordered sequences of amino acids has been identified experimentally. This unique answer is the process of self-sequencing (self-ordering; self-instructing) of amino acids[8]. Moreover, the emergence of ordered sequences within polyamino acids from matrices of free amino acids comports with fundamental tenets of evolutionary theory[9].

The total evidence for self-ordering is both direct and indirect. Direct evidence concerns comparisons of total compositions with those of terminal residues, for several decades[10–12] a kind of analysis most easily carried out. The most direct evidence is the kind most recently obtained[13]. The latter is based on determination of sequences in peptides produced by thermal copolymerization of amino acids and qualitative and quantitative comparison with what was predicted on the assumption of randomness.

The earliest and yet most voluminous of the evidence is the indirect kind. This evidence results from the fact that, like other ordered molecules, self-ordered macromolecules are much alike from one molecule to another. When they are much alike, heterogeneity is constrained, much as in modern organisms the molecules are constrained in heterogeneity by the operation of the genetic code. It is such evidence for self-instruction that permitted proceeding with the construction of a theory of protobiogenesis. The derived inference was that the first (thermal) proteins did not need instructions from a prior DNA; they obtained such information instead from the reactant amino acids. The first informational macromolecules were accordingly proteins.

The evidence for self-ordering is extensive enough to defy comprehensive review[14]. Some of the indications will however be reviewed here.

COMPOSITION

The first test for randomness by the polymer chemist is that of analytical comparison of the monomer feed stock with the composition of the polymer. If the

two are the same, the copolymer is random, or "ideal"[15]. Accurate studies of amino acid composition became feasible when columnar methods of fractionation were introduced[16] into laboratory use. In each case examined, the composition of the amino acid mixture differed from that of the thermal copolymer[17].

Compositions of proteinoid remained constant when the polymer was purified and repurified by solution in hot water and cooling[18]. An equal recovery of fractions of dissimilar compositions from mixtures on each purification is theoretically not excluded, but it is unlikely through two purifications.

Disparity in analyses of total composition and terminal residues

If a given polyamino acid were randomly constituted, its total composition should theoretically be the same in all positions in the peptide chains. A number of analyses have shown, instead, disparities between C-terminal and N-terminal analyses, and between terminal and total analyses[14]. The polymers are therefore non-random.

The extension of this result is that in which two tripeptide sequences have been identified in one major fraction of a thermal copolyamino acid[13].

Fractionation of a proteinoid on Sephadex and other columns

In a first thorough examination of the heterogeneity of a proteinoid, one such preparation was first separated into components on DEAE-cellulose[19]. Six major fractions were eluted. The limited number of fractions was a strong first indication of non-random composition.

Individual fractions Nos. 3, 4, and 5 were analyzed, found to be compositionally quite similar, close to stochiometric in amino acid ratios, and quite similar to the unfractionated polymer in composition. Final purification of the fractions on Sephadex G-25 and G-50 gave symmetrical peaks by removal of small tails, and provided estimates of molecular weight in the range of 4000–6000 by sedimentation.

The thus-purified fractions were reaffirmed as showing only sharply constrained heterogeneity, by methods such as the sedimentation analysis, peptide mapping of partial hydrolyzates, and electrophoresis.

This was the first study that thoroughly established the constrained heterogeneity of thermal copolyamino acids. Numerous other studies in other laboratories have extended the observations, which essentially make the same point that amino acids order themselves upon thermal copolymerization[8].

Self-ordering of amino acids and historical thought

The proteinoid model[8] is the result of experimentally retracing primordial steps in molecular evolution. Analysis of modern cells cannot be expected to provide data on how assembly of primordial cells occurred[20]. However, questions of later evolution require assumptions about the course of earlier evolution. As a consequence of the urgency of such questions, a body of assumptions has grown. When data became available from the heuristic retracement of molecular evolution, some long-standing premises and descriptive terms were found to be contravened[6]. Some discussion of the older ideas is desirable here as a prelude to explanation of the perceived significance of self-sequencing.

Outstanding among the older premises are those of: an initial random matrix,

the term "self-replication of DNA", and the idea that genes "make" enzymes.

While the concept of randomness in evolution has occasionally been qualified[21,22], it has also been analyzed as a statistical concept in a biological context. Eden's treatment of random variation has indicated that such evolution is highly implausible without determinate features[22]. The widespreadness of the concept of random matrix (e.g., ref. 3) is basically incompatible with the non-randomness experimentally identified in the self-sequencing of amino acids[8]; such self-sequencing appears to be a determinate feature.

Terms such as the "self-replication of DNA" are used by authors whose other writings indicate that they are actually not thinking that DNA actively replicates itself but, rather, that DNA *is* replicated, the latter statement being the more precise of the two. Enzymes and additional proteins are required[23] for the replication. "Self-replication of nucleic acid" is, however, a misleading[7] but frequently encountered statement. In addition, DNA is not directly replicated as such; it is replicated through two complementary syntheses[24].

The sequential circumstances of modern replication → transcription → translation are not necessarily transposable to primordial events as stated by Crick in 1970, in accord with a quotation earlier in this paper.

The seminal participation by proteins in the protobiological process of molecular reproduction is consistent with the significance of the principle of self-sequencing of amino acids, since that principle yields proteins having non-random specificity. This new view has been regarded as in conflict with the idea of "self-replication of DNA".

The historical view on DNA's being replicated by proteins (and ATP) instead of replicating itself, has been recently stated by Kornberg[25]: "Although the Watson–Crick proposal for the replication of DNA had not predicted the operation of an enzyme, the properties of DNA polymerase suited the role of accurate chain assembly."

Kornberg's emphasis on proteins is further supported by the large number of specific enzymes involved in the replication process[7,23,26]. (Here too, chromatographic fractionation as on Sephadex has been an essential tool). The persistence of the assumption that DNA, as genes, makes macromolecules such as enzymes can be found in another historically relevant reference published in 1976[27].

SIGNIFICANCE OF SELF-ORDERING OF AMINO ACIDS

The positive significance to life and evolution of the self-ordering of amino acids is basically rooted in thinking that led to questioning assumptions easily drawn from analysis of modern organisms. Several of the positive contributions to thinking have been described[8]; others are being systematized[28].

Two or three may be mentioned here. One is the development of a theory of protobiogenesis arising out of simulating the archaic automatic origin of protolife in the laboratory; the other is the supplementation of Darwin's theory of natural selection by an improved definition of the limits of natural variation.

The chicken–egg question was formulated by biochemists. Its resolution required protobiochemistry. The theory not merely merited confirmation by experiment in the usual way[20], it required experiments to find an answer.

Dickerson[29] has recently emphasized the fact that the theory of the origin of life is relatively complete, and that this aspect is of more moment than the less abstract event of producing a living organism in the laboratory[29]. These evaluations constitute progress from the assessments of Pirie who wrote on "The meaninglessness of the terms life and living"[30], and who seventeen years later published a paper, "On making and recognizing life"[31]. This second paper appeared in 1954, almost the same year as the published report on the double helix[32], which shifted attention from life-cycle protein to inheritance-related nucleic acid.

Pirie stated that, in his personal view, "biopoesis" would be recognizable by the criteria of liquid content of an eobiont, its operation below 200°C, its aseptic production, catalysis of at least five or six reactions, synthesis of the catalysts themselves, and the ability to reproduce. With recent findings that organisms containing lysine-rich proteinoids can catalyze the synthesis of peptides from any kind of amino acid[33,34], these requirements are met[14,34]. Of course, further refinement is to be sought; reproduction as the direct result of cellular peptides making peptides, instead of reproduction occurring only at the higher hierarchical level of cellular proliferation, would be a step forward toward a more modern organism.

While Pirie recognized a gradual nature to evolution, he did not perceive a stepwise evolution[14], and he missed especially what we now recognize as the crucial aspect of self-ordering[8]. This latter was made vivid by experiments which in turn, as stated, emerged especially from chromatographic studies of fractionation[1] and analysis[16].

Awareness of the stepwise nature of protocell genesis and of the self-ordering prelude to it[14] spelled the large difference between whether the proteinoid microsphere was a first, and perhaps only, answer to Pirie's question, or whether it was some kind of incompletely defined laboratory curiosity.

The relationship of studies on proteinoid to Darwinian evolution resides in the question of the relationship of self-ordering to natural selection. Darwin[36], and subsequently Morgan[37], pointed out that evolutionary theory needed an explanation for natural variation upon which natural selection could act. Eden's analysis of the popular neo-Darwinist view[22] is a modernized expression of that need. Self-ordering qualitatively reveals that natural variation is and was so limited that evolution as we know it could indeed have proceeded as it has.

This narrow evolutionary highway is a modern view, in essence one that was not possible before the second half of this century. It is this kind of significance that owes its recognition to the contributions of pioneers in methods of chromatographic fractionation of copolyamino acids. Effective methods of fractionation have made a theory of protobiochemistry possible; the results have established protobiochemistry as a comprehensive chapter in biochemistry[35].

Calvin[38] and Dillon[39] have each proposed that modern organisms contain relicts of the archaic self-ordering of amino acids. Since amino acid side chains are primary components of specific interactions and reactions[14], I propose that amino acid self-ordering evolved directly to the specificities of enzyme–substrate reactions, both kinds of specificity being rooted in the arrangements of the same amino acid side chains. In this and other ways we visualize that protobiochemical phenomena evolved to biochemical phenomena[28].

ACKNOWLEDGEMENT

The experiments and theory from which this paper has emerged have been mainly supported by Grant No. NGR-10-007-008 of the National Aeronautics and Space Administration. Contribution No. 343 of the Institute for Molecular and Cellular Evolution.

REFERENCES

1 J. Porath, *Trends Biochem. Sci.*, 3 (1978) N100.
2 K. Popper, in F. J. Ayala and T. Dobzhansky (Editors), *Studies in the Philosophy of Biology*, Macmillan, New York, 1974, p. 259.
3 J. Monod, *Chance and Necessity*, Alfred A. Knopf, New York, 1971, p. 143.
4 F. H. C. Crick, *Nature (London)*, 227 (1970) 561.
5 J. Monod, *Chance and Necessity*, Alfred A. Knopf, New York, 1971, p. 98.
6 S. W. Fox, *Naturwissenschaften*, 60 (1973) 359.
7 L. S. Dillon, *The Genetic Mechanism and the Origin of Life*, Plenum Press, New York, 1978, p. 204.
8 S. W. Fox, in W. H. Heidcamp (Editor), *The Nature of Life*, University Park Press, Baltimore, MD, 1978, p. 23.
9 T. H. Morgan, *The Scientific Basis of Evolution*, W. W. Norton & Co., New York, 1932, p. 218.
10 S. W. Fox and K. Harada, *J. Amer. Chem. Soc.*, 82 (1960) 3745.
11 P. Melius and J. Y.-P. Sheng, *Bioorg. Chem.*, 4 (1975) 385.
12 J. Rosmus and Z. Deyl, *Chromatogr. Rev.*, 13 (1971) 163.
13 T. Nakashima, J. R. Jungck, S. W. Fox, E. Lederer and B. C. Das, *Int. J. Quant. Chem.*, QBS4 (1977) 65.
14 S. W. Fox, *Naturwissenschaften*, 67 (1980) 378, 576.
15 F. W. Billmeyer, Jr., *Textbook of Polymer Science*, Interscience, New York, 1962, p. 312.
16 S. Moore, D. Spackman and W. H. Stein, *Anal. Chem.*, 30 (1958) 1185.
17 P. Melius, in E. E. van Tamelen (Editor), *Bioorganic Chemistry*, Vol. III, Academic Press, New York, 1977, p. 123.
18 S. W. Fox, K. Harada, K. R. Woods and C. R. Windsor, *Arch. Biochem. Biophys.*, 102 (1963) 439.
19 S. W. Fox and T. Nakashima, *Biochim. Biophys. Acta*, 140 (1967) 155.
20 S. W. Fox, in E. E. van Tamelen (Editor), *Bioorganic Chemistry*, Vol. III, Academic Press, New York, 1977, p. 21.
21 S. Wright, in P. S. Moorhead and M. M. Kaplan (Editors), *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, Wistar Institute Press, Philadelphia, PA, 1967, p. 117.
22 M. Eden, in P. S. Moorhead and M. M. Kaplan (Editors), *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution*, Wistar Institute Press, Philadelphia, PA, 1967, p. 5.
23 A. Kornberg, *DNA Synthesis*, W. H. Freeman & Co., San Francisco, CA, 1974, pp. 26, 58*ff*.
24 J. D. Watson, *Molecular Biology of the Gene*, W. A. Benjamin, Menlo Park, CA, 3rd ed., 1976, p. 203.
25 A. Kornberg, in A. Kornberg, B. L. Horecker, L. Cornudella and J. Oro (Editors), *Reflections on Biochemistry*, Pergamon Press, New York, 1976, p. 248.
26 U. Lagerkvist, *Amer. Sci.*, 68 (1980) 192.
27 I. Shine and S. Wrobel, *Thomas Hunt Morgan*, University of Kentucky Press, Lexington, KY, 1976, p. 142.
28 S. W. Fox, in preparation.
29 R. E. Dickerson, *Sci. American*, 239, No. 3 (1978) 70.
30 N. W. Pirie, in J. Needham and D. E. Green (Editors), *Perspectives in Biochemistry*, Cambridge University Press, Cambridge, 1937, p. 11.
31 N. W. Pirie, *New Biol.*, 16 (1954) 41.
32 J. B. Watson and F. H. C. Crick, *Nature (London)*, 171 (1953) 736.
33 T. Nakashima and S. W. Fox, *J. Molec. Evol.*, 15 (1980) 161.
34 S. W. Fox and T. Nakashima, *BioSystems*, 12 (1980) 155.
35 A. L. Lehninger, *Biochemistry*, Worth & Co., New York, 2nd ed., 1975, p. 1048.
36 C. Darwin, *The Origin of Species*, Random House, New York, undated, p. 372.
37 T. H. Morgan, *The Scientific Basis of Evolution*, W. W. Norton & Co., New York, 1932, p. 238.
38 M. Calvin, *Chemical Evolution*, Oxford University Press, Oxford, 1969, p. 169.
39 L. S. Dillon, *The Genetic Mechanism and the Origin of Life*, Plenum Press, New York, 1978, p. 209.