SILVANO P. COLOMBANO

TOWARDS A THEORY OF THE ORIGIN OF THE GENETIC CODE: THE COMPLEMENTARY ROLES OF CHEMICAL AND LOGICAL CONSTRAINTS

February 1977

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Department of Biophysical Sciences of the State University of New York at Buffalo.

ABSTRACT

The results of prebiotic chemistry and the main schools of thought on the origin of the genetic code are reviewed, with particular emphasis on the conflict between the stereochemical and the stochastic approach to the problem.

The assumptions underlying Eigen's theory of self-organization of matter are also discussed.

A model is proposed for a stage of evolution, preceding the onset of the genetic code, during which adaptors for all possible codon-amino acid assignments (belonging to different codes) could be present. Heuristic computations based on a very simple version of the model, point to a natural tendency for the system to eliminate competition and allow for the survival of only one self-consistent (self coding) set of adaptors; i.e. one code is chosen.

Although mainly stochastic, the model allows in principle for hypothetical stereochemical affinities between anticodons and amino acids. These would favour the choice of one particular code, but it appears unlikely that all final anticodon - amino acid assignments could ever be due to such affinities. Some incompatibilities with Eigen's theory are pointed out.

Suggested for further study are more realistic versions of the model, which would include the possibility of varying

the rate of self-replication and mutation of the genes, and the inclusion of adaptors with different discriminating abilities. It is proposed that the genetic code evolved through stages in which only classes of amino acids were distinguished, in agreement with a model already proposed by Woese.

ACKNOWLEDGMENTS

I consider myself very fortunate for having had the opportunity to interact with Dr. Howard H. Pattee, and to enjoy his friendship. He has had a profound impact on my scientific and philosophical thinking, although, unfortunately my work seldom manages to bring his positive influence to light. I also wish to thank him, and Mrs Mary Ellen Pattee, for graciously allowing me to spend several "work weekends" at their home in Binghamton.

I would like to acknowledge the contribution of Vahe Bedian in helping me develop and clarify many of the concepts expressed in this work.

Finally, I wish to express my gratitude to Drs. Fred Snell, Robert Rein and Robert Spangler, for helping me in different ways on several occasions, and to the faculty and students of the Department of Biophysical Sciences for always providing a pleasant and supportive environment.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

The discovery and "cracking" of the genetic code can already be hailed as some of the great achievements of modern science. Yet, interestingly, no consensus has been reached on its fundamental nature. The following quotations by some of th leading researchers in the field illustrate the ongoing contr

Prof. Sidney Fox (1974b):

> The results are consistent with a stereochemical basis for the genetic code rather than a "frozen accident" of statistical occurrences.

Prof. Geoffrey Hoffmann (1975):

> A... school of thought (11,18) considers that a systematic evolutionary development began only with the nucleation of an autocatalytic cycle.... The model seems, however, to imply that a great deal happened all at once.... It will be seen that such a nucleation is perhaps more probable than would appear at first sight.

---

11. Eigen, M. 1971. Naturwissenschaften 58: 465-523; Quart. Rev. Biophys. 4:149

18. Hoffman, G.W. 1974. J. Mol. Biol. 68: 349-362

Prof. Melvin Calvin (1975):

> I think this code arose not by accident (33,34, 35) but because of the peculiar chemistries of the various bases and amino acids.

---

33. M. Eigen 1971. Naturwissenschaften 58:465

34. F.H.C. Crick 1968. J. Mol. Biol. 38:267

35. L.E. Orgel 1968. J. Mol. Biol. 38:381

Prof. Thomas Jukes (1974):

> The proposal that the genetic code originated by some kind of loose affinity between the amino acids and the bases is an obvious one....
>
> An objection to this line of reasoning is that it tends to invoke the existence of interactions that are so weak that they cannot be detected except under specialized and artificial conditions. I feel that such an explanation may rely on a desire to postulate non-existent phenomena because we feel intuitively that they ought to have existed since we cannot think of any other way that the genetic coding of proteins might have started [italics mine].

The major motivation for this work stems from the realizatic that the problem of the origin of the genetic code presents a challenge to intuition even before it can be formulated. The exis of a genetic system gives us a clear understanding of Darwin's theory of evolution: mutations in the gene cause changes in the phenotype; if these changes are advantageous to the survival of the organism, survival and reproduction cause the mutation to be retained, and eventually to be transmitted to all members of the same species. The clarity of this picture however, perhaps by cont makes it impossible to understand how the genetic system itself evolved. As understood above, the keys to evolution are:

1. gene replication,which propagates the mutation;
2. a consistent relationship between the genotype and the phenotyp as provided by the genetic code.

In other words, the existence of the genetic system is needed for evolution. How are we then to understand the evolution of the gentic system itself? Was it then simply an (unlikely or likely) accidental event? While on one hand it is difficult to conceive of evolution in the absence of a code, it is also difficult to abandon the idea that life was the product of a continuous, inevit process.

The challenge we are faced with is to understand the evolutio of matter before , and leading to, the onset of the genetic code. Clearly such understanding would have profound philosophical implications for our concept of life, and consequently of ourselves

From the point of view of science, a theory of the evolution of matter in the absence of a code implies the ability to predict self-constraining behaviour in complex chemical systems. At least some results could probably be extrapolated to all complex open systems and thus have far reaching consequences. In fact, efforts in this general direction have already been under way from the po of view of irreversible thermodynamics, independently from the problem of the origin of life. The work of Glansdorff and Prigogi (e.g. 1971) on symmetry breaking instabilities, is representative of this more general approach.

Unfortunately the scientific relevance of the problem of the origin of life (or, more specifically, the origin of the genetic code) is not matched by its amenability to scientific investigatic One deals with one or more, probably unique,occurrences, and with time spans of billions of years. Thus no explicit tests would ever be possible for any theory which tried to encompass the whole process. One can only select as many relevant phenomena as possibl and devise models which extrapolate between them.

This work is unique in its emphasis on the following points:

1. At least three phenomena can be studied separately. They are a) non enzymatic template replication of polynucleotides, b) evolution of a polypeptide constructing machinery in the absence of the genetic code, c) evolution of a translating machinery (onset of the genetic code).
2. Focusing of phenomenon c), I argue that the code arises from a complementary relationship between chemical and logical constrai Any stochastic or deterministic approach which ignores either type of constraint cannot form the basis of a valid theory.

3. The idea that the genetic code resulted from a continuous inevitable process is not evident in my approach. However, if one allows that enzyme function evolved from recognition of less detailed, to recognition of more detailed propertie it becomes possible to envision successive primitiv codes, and the continuity problem is reduced to that of understanding the transitions between these successive code The outcome of this analysis is a model which is essentiall in agreement with the Translation Error model, proposed by Woese in 1967, but shows more clearly the kinds of proble a theory would have to solve.

The essential chemical constraints I consider are: a) exis of a polypeptide constructing machinery in the absence of the code, b) folding of polypeptide sequences to form three-dimensi structures which can perform functions (I am particularly inter in codon-amino acid adaptor functions), c) enzyme-free replicat of polynucleotides, d) affinities (of unspecified nature) betwe codons and amino acids.

The idea of logical constraint is best explained after the examples of chapter IV, but it can be introduced here with an analogy. There are mophological and syntactical elements which are common to all natural languages, such as the existence of nouns, verbs, modifiers, onomatopoeic sounds. Presumably these elements are strictly related to human perception and physiology (e.g. ability to produce certain types of sounds). If one views the genetic code as a biological language, the above elements are comparable to chemical constraints, which would certainly

be common to any code one might envision. The question is: would different codes be possible, the same way different languages are possible? Or, more precisely, would it be possib to have alternative sets of codon-amino acid assignments the same way we have different dictionaries? If, as suggested by Pattee (section III.1.3), the answer is yes, the particular choice of a code can be viewed as a logical constraint. This emphasises the idea that an alternative choice was possible, while the same could not be said for chemical constraints. As a reminder of the analogy, logical constraints will sometim be called "linguistic".

I mentioned earlier that it is difficult both to envision continuity and chance events for the origin of the genetic code In working towards a picture which could overcome both difficul I tried to develop models which were suitable for heuristic computation, and could possible be the starting point for a mor general mathematical theory of self-organization.

Chapter II and III are reviews of relevant experiments in prebiotic chemistry and of ideas concerning the origin of the gentic code, respectively. Chapter IV introduces the fluctuatio model, the concept of self-coding and some simple computer gam These are designed to illustrate competition between codes, and the interaction between chemical and linguistic aspecs of codin In chapter V the fluctuation model is evaluated in light of the problem of the evolution of the code itself, and an evolutionar model is proposed. This model is hardly more than a sketch, yet in my opinion, it illustrates the most promising approach to the problem of the origin of the genetic code.

II. EXPERIMENTAL DATA RELEVANT TO THE MODELING OF PREBIOTIC SYSTEMS

II.1 The Primitive Earth

The final consolidation of the Earth's surface occurred 4 to 5 billion years ago. The most ancient traces of life are reported to be at least 3.1 billion years old (Kenyon and Steinman 1969, p. 76). There is no general agreement as to what particular type of environment was most likely to favour the early stages of molecular evolution; prebiotic conditions are generally postulated on the basis of some compromise between geological evidence on the conditions prevailing on the primitive Earth, and present day biochemical know-how. The origin of life may have been a wide spread phenomenon, or it may have been possible only because of some peculiar set of local circumstances, likely only in few regions of earth. This difference alone may be a significant aspect in choosing models of the origin of life, but since no evidence exists one way or the other, I shall not pursue this difference. I shall present dat on prebiotic synthesis which I believe to be relevant to a discussion on the origin of the genetic code, but I have not attempted to follow a single set of assumptions on prebiotic conditions, nor to evaluate the results in any way. Hopefully it will become clear that my approach to the problem of the origin of the genetic code requires only very general biochemica data.

The primitive Earth could have offered a wide variety of environments and energy sources: seashores, river deltas,

volcanoes, thermal springs, contact with special minerals and with ice; thus possible exposure to different degrees of solar radiation and thermal energy were possible. Even conditions created by the impact of meteorites have been considered.

The primitive atmosphere of the Earth was probably composed of some mixture of CO_2, CH_4, NH_3, N_2, H_2O vapor, H_2S and H_2 (Lehninger 1970, p. 770). The exact composition is uncertain but it is generally agreed that oxygen was only a trace constituent (Ibid., p. 771). As a consequence the ozone layer was probably missing and UV radiation could penetrate the troposphere and provide an abundant source of free energy; electrical discharges were another possible source. As far as the hydrosphere is concerned there is disagreement on how fast the volume of water increased, but the inorganic composition of the oceans is believed to have remained essentially unchanged (Keny and Steinman 1969, p. 113). Commenting on how the wide range o experimental conditions allowed by this picture of the primitiv Earth still yields fairly consistent patterns of chemical evolu tionary reactions, the same authors (Ibid., p. 118) come to the "tantalizing conclusion" that:

> ...knowledge of the details of Earth history, which will probably remain hidden from us in any case, may not be required for a complete understanding of biochemical origins.

From the point of view of a study of the origin of the code this gives one more reason not to be too selective too soon in considering prebiotic conditions: it is natural to expect that conditions which favoured the production of the "building blocks

of biological systems would also favour higher level of organi zation such as the genetic code, but this is not necessarily t case. If the optimal environmental conditions are different f the two cases then intermediate conditions must be found, or a mechanism for the needed transition. In the meanwhile it is certainly wise to pursue all reasonable directions and startin points, not only on the basis of the good results they may yie but because there is a chance that the pathway (or pathways) which leads from simple organic molecules to complex biologica systems will become evident only from a broad, unselective picture of possible prebiotic conditions.

II.2 The "Building Blocks"

II.2.1 Synthesis of Amino Acids

Miller (1953) reported the production of substantial yield of organic molecules and amino acids from a postulated primitiv earth atmosphere enclosed in flasks heated for periods of one to two weeks at 100°C. In another series of experiments (Mille 1955, 1957a,b, 1959) methane, ammonia, water and hydrogen, when subjected to electrical discharge, produced Glu, Asp, Ala, Ser and Gly. The same amino acids resulted from thermal treatment of a primitive atmosphere (Harada and Fox 1964) in a way which

was sequentially compatible with successful experiments on pol
merization of amino acids (Fox and Harada 1960a,b). Aspartic
acid and asparagine (and cytosine) were also obtained from sim
reactions of cyanoacetylene with inorganic substances in aquec
solution (Sanchez et al. 1966); cyanoacetylene is readily pro-
duced by electric discharge on a mixture of methane and nitrog
Aldehydes have then been suggested as possible prebiotic react
formaldehyde, heated with ammonia, yields a product which can
hydrolized to the same amino acids mentioned above and, in add
tion, valine and proline (Fox and Windsor 1970). Subbaraman e
al. (1972) obtained similar results by diluting glycine with
formaldehyde in a mildly alkaline solution. In the same exper
ment threonine and β-phenylserine were also detected when form
dehyde was substituted with acetaldehyde and benzaldehyde. UV
radiation on a fog of formaldehyde and ammonium nitrate caused
the formation of glycine, alanine and possibly threonine
(Pavlovskaya et al. 1971). A primitive atmosphere which inclu
H_2S , when exposed to the same radiation, produced Ala, Gly, S
Glu, Asp and Cys (Ito and Bowman 1971). More recently water-
surface quench reactions have been studied by Ponnamperuma's
group: one of their effects is to double the production of am
acids (Park et al. 1975).

Experiments on abiotic synthesis of amino acids from post
lated primitive atmospheres will certainly continue for a long

time, as researchers strive to find a most plausible chem
cal pathway which can gain general acceptance. At any rate,
laboratory results and recent evidence from exobiology (which
will outline below) have convinced most workers that the most
common amino acids or their precursors were readily available
the primitive Earth.

One objection, however, has been raised by Cairns-Smith (1975) to the common interpretation of the results I have just sketched and in general to all the results of prebiotic chemist

> The main conclusion from abiotic syntheses is that modern organisms prefer molecular units that are easy to make. This could be the result of selection pressure during very early evolution rather than a reflection of the large scale molecular environment within which life first arose.

In this work I assume that whatever is "easy to make" is at lea
tried first but it is not necessarily maintained within a syste
Understanding the conditions under which this may happen is a
major aim of my study.

II.2.2 Synthesis of Nucleotides

Bases

The pyrimidine base which is most easily produced under prebiotic conditions is uracil. Fox and Harada (1961) synthesized it by heating malic acid, urea and polyphosphoric acid

between 100° and 140°C. Heat at 130°C was also used by Orò (19
to obtain uracil from urea and vinyl cyanide in an aqueous solu
tion of ammonia. Cytosine was synthesized by Orgel's group
(Ferris et al. 1968) from cyanoacetylene and cyanate; cyanoace-
tylene was produced by electric discharge on nitrogen gas and
methane. The same authors feel that the relative instability o
these two compounds would not allow for a wide range of prebiot
environments. Cytosine itself can yield uracil, by hydrolysis
(ibid), and with the help of ionizing radiations (Ponnamperuma
et al. 1962). Sherwood et al. (1971) suggested methylation of
uracil with formaldehyde and hydrazine as a possible prebiotic
pathway for the synthesis of thyamine.

Of the purine bases adenine is found most readily. Heating
a concentrated solution of ammonium cyanide gave successful re-
sults (Orò and Kimball 1960,1962, Orò 1961). Adenine was also
found when a mixture of methane, ammonia, water and hydrogen wa
irradiated with an electron beam (Ponnamperuma et al. 1963).
Ponnamperuma and co-workers (1964) considered the possibility o
producing purine and pyrimidine bases by heating a mixture of
amino acids: in such an experiment they detected guanine. San-
chez, Ferris and Orgel (1967,1968) showed that an aqueous solu-
tion of hydrogen cyanide would allow for reaction pathways lead
to the formation of adenine, guanine and hypoxantine. They sug-
gested that the needed concentration of HCN could be achieved by
cooling to separate out ice (id. 1966, 1967).

Sugars

It has been known since the last century that formaldehy in an aqueous alkaline solution could form a mixture of mono- saccharides. Objections that the alkalinity required was too high (terms of prebiotic conditions) led investigators to study the possibility that a base catalyst might not be needed in the presence of U radiation: this type of experiment was performed successfull by Ponnamperuma and Mariner (1963 and Ponnamperuma 1965), who reported the formation of ribose and deoxiribose. In the sea for possible prebiotic catalysts, experimenters have been t ing to clays which, as we shall see, may have played a very i portant role in many types of reactions. Their role in the synthesis of sugars has been studied by Gabel and Ponnamperum (1967) who used alumina, kaolite and illite refluxed with an aqueous solution of formaldehyde (possibly a model for a ther spring): formaldehyde was converted to monosaccharides.

However there is no general agreement on formaldehyde as a precursor of sugars. Horowitz and Miller (1962) objected that the concentration needed was unrealistic for prebiotic c ditions; Reid and Orgel (1967) reported similar results to th of Gabel and Ponnamperuma (1967), nevertheless they felt the above objection was still valid; they added further that the stability of the sugars formed was also a problem.

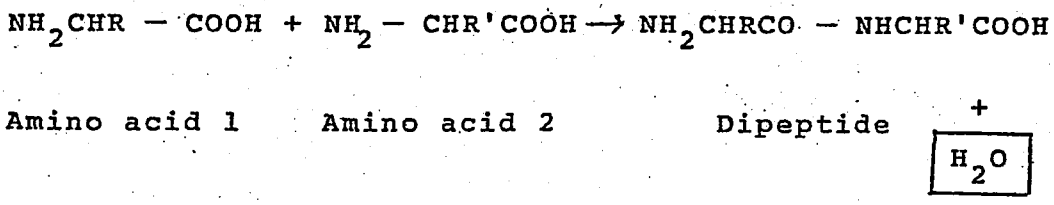
Condensation reactions: synthesis and phosphorilation of nucleosides.

The condensation of most biologically important monomers and polymerization, involves dehydration reactions (fig. II-1) This presents a problem since the presence of water is postula for most prebiotic environments, and the needed covalent bonds undergo hydrolysis rather easily. Two factors can be controll to favour dehydration: temperature and the presence of appropr ate condensing agents. In addition UV radiation may have play a role.

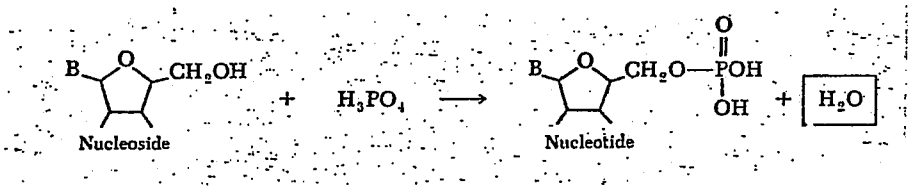
Ponnamperuma and co-workers (1963a,b) reported the format of adenosine, AMP, ADP and ATP in a solution of adenine and ri bose in the presence of phosphoric acid and UV radiation. Steinman, Lemmon and Calvin (1964) confirmed the same finding and suggested further that cyanamide might have played a key r in chemical evolution. Adenine was linked to deoxyribose in t presence of hydrogen cyanide; UV light enhanced the production deoxyadenosine but was not essential (Ponnamperuma and Kirk 19 More recently Sanchez and Orgel (1970) proposed that reactions of sugars with cyanamide and cyanoacetylene could form a pre- biotic route to cytidine and uridine; UV light was also needed for anomerization and epimerization. In their words this route would be "inefficient, but most plausible" of those reported to that date. Nucleotides have been formed by heating the nucleo- sides with inorganic phosphates (Ponnamperuma and Mack 1965, Beck and Orgel 1967, Lohrmann and Orgel 1968, Chang et al. 1970

FIGURE II-1

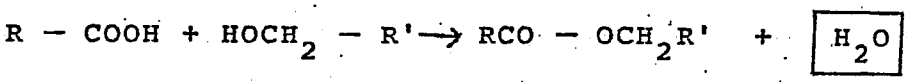
Examples of the condensation of biologically important monomers. After Kenyon and Steinman 1969, p. 164.



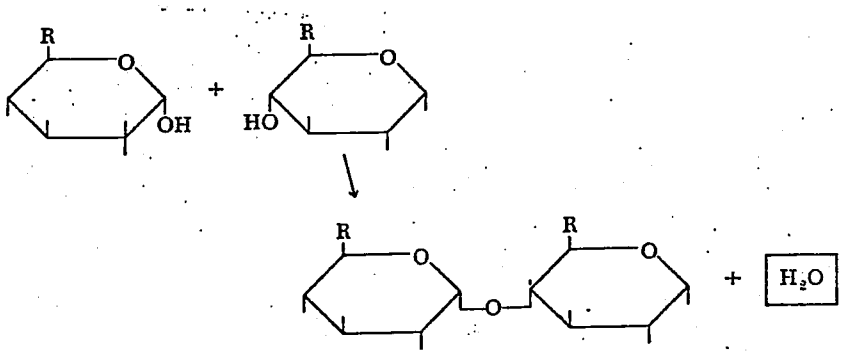
(b) Phosphate ester (B = heterocyclic base):[1]



(c) Carboxylic ester (= lipid, if the alcohol is glycerol):



(d) Glycoside:[1]



[1] Branched bonds represent hydroxyl groups

or at temperatures from 0° to 22°C with polyphosphoric acid (Waehneldt and Fox 1967). Lohrmann and Orgel (1971) pointed o that phosphorylation of nucleosides had succeeded only with ac phosphates, they suggested therefore the addition of urea and ammonium chloride which allow phosphorylation with neutral or basic phosphates. Such a mixture, with thyamidine and Na_2HPO_4 yielded Tp, pT and pTp (Bishop et al. 1972). With urea higher phosphorilation products are also obtained from TMP (Odon et a 1973), and from UMP in the presence of struvite ($MgNH_4PO_4 \cdot 6H_2O$ which, according to Handschuh and Orgel (1973), could have pre cipitated from evaporating seawater. However, in the presence of apatite, which according to Orò and Stephen-Scherwood (1974 was the most probable source of phosphate on primitive Earth, cyanide and dicyanamide proved to be more effective than urea for the phosphorilation of uridine (Schwartz 1972).

II.2.3 Extraterrestrial Evidence

Perhaps some of the most convincing evidence that the buil ing blocks of living systems could have been formed under a ver broad range of conditions, is the presence of organic material in extraterrestrial space. Several molecules of relevance to prebiotic chemistry have been identified in interstellar space (table II-1), in meteorites (Urey 1966, Kvenvolden et al. 1970,

TABLE II-1

Molecules identified in interstellar space.
(After Herbig (1974))

	CH	Silicon monoxide	SiO
	CN	Acetonitrile	CH_3CH
	CH^+	Carbonyl sulfide	OCS
Hydroxyl	OH	Isocyanic acid	HNCO
Ammonia	NH_3	Methyl acetylene	CH_3C_2H
Water	H_2O	Hydrogen isocyanide	HNC?
Formaldehyde	HCHO	Formamide	NH_2HCO
Carbon monoxide	CO	Thioformaldehyde	H_2CS
Hydrogen	H_2	Acetaldehyde	CH_3CHO
Hydrogen cyanide	HCN	Formaldimine	CH_2NH
	HCO^+?	Hydrogen sulfide	H_2S
Cyanoacetylene	HC_3N	Carbon monosulfide	CS
Methyl alcohol	CH_3OH	Sulfur monoxide	SO
Formic acid	CHOOH		

? Tentative

1971, Lawless et al. 1971, Kvenvolden 1974), and on lunar sam
(Modzeleski et al. 1973, Fox et al. 1973). Of course great c
has been taken to minimize and estimate the effects of contam
nation from Earth, whenever samples have been studied, and to
establish beyond doubt the extraterrestrial origin of the com
pounds.

It is particularly interesting to notice that the amino acids which are most abundant in animal proteins are also read
ly produced in prebiotic experiments, and most common in extra
terrestrial space (table II-2). Fox (1975) remarks:

> This is particularly gratifying to a biochemist who has watched as the evidence accumulated over several decades for the unity of biochemistry... Now it appears that this unity of biochemistry had its origins in a kind of common cosmochemistry... We live in an internally ordered universe .

The possibility has also been mentioned that "d
rected panspermia" in fact brought life to this planet from ot
sites (Crick and Orgel 1973). In this case experiments based
primitive Earth conditions might even be misleading, neverthe
the fundamental questions on life as a possible (inevitable, p
bable, improbable?) physical phenomenon would still not be answered

TABLE II-2

Typical examples of amino acids from various sources. After Fox (1975)

Lunar Samples	Meteorites (Chondrites)	Terrestrial Lava (Exterior of Sample Heated)	Chemical Synthesis (from Formaldehyde & Ammonia)	Animal Proteins
Glycine	Glycine	Glycine	Glycine	Aspartic Acid
Alanine	Alanine	Alanine	Alanine	Glutamic Acid
Glutamic Acid	Glutamic Acid	Glutamic Acid	Glutamic Acid	Glycine
Serine	Valine	Serine	Serine	Alanine
Threonine	Proline	Threonine	Valine	Leucine
		Isoleucine	Proline	Isoleucine
		Leucine		Proline
		Valine		Valine
				Lysine
				Threonine
				Arginine
				Cystine
				Phenylalanine
				Histidine
				Tryptophan
				Methionine

II.3 Polymerization[1]

II.3.1 Synthesis of Oligo- and Polypeptides

Polypeptides have been prepared by heating mixtures of acidic or basic amino acids in the presence of phosphoric or polyphosphoric acid (Fox and Harada 1960a,b, Harada and Fox 19 Fox and Nakashima 1967). A certain lack of randomness in the order of residues was reported very soon (Fox and Harada 1960b and later confirmed in the same and in other laboratories (see below). The range of compositions of amino acids which could be condensed thermally was found greater than expected at firs (Fox and Waehneldt 1968), and the properties of the resulting "proteinoids" became, of their own right, object of research. I will report on them in a separate section to follow.

Synthesis mediated mainly by condensing agents has only produced oligopeptides so far: cyanamide in aqueous solution of glycine and leucine exposed to UV light formed some dipeptides and one tripeptide (Ponnamperuma and Petersonn 1965); polymers of glycine at least up to the tetrapeptide were obtai by Calvin's group with dicyanamide (Steinman et al. 1965,1966). More recently diaminomaleonitrile (tetramer of HCN) and aminoacetonitrile have been used by Ponnamperuma's group (Chang et al. 1969, Chada et al. 1971) to produce dipeptides.

The possible role of clays and crystals in polymerization is also being studied. Paecht-Horowitz, Berger and Katchalsky (1970, Paecht-Horowitz 1972) used amino acyl adenylates (at present the active form of amino acids) concentrated by adsorpt

---

[1] See also condensation reactions, section II.2.2.

on montmorillonite, a naturally occurring clay, to produce poly peptides of up to 40 units. Burton and Neuman (1971) proposed that hydroxyapatite, a crystal, could function as catalyst to allow for the polymerization of amino acids at a lower temperature than that normally required; they tried this successfully with glycine.

Although no plausible prebiotic pathway for the production of the adenylates of amino acids has been found (Hoffman 1975), they can serve as a model to study the control of amino acid sequence in the absence of highly evolved cellular constraints (Krampitz and Fox 1969, Banda and Ponnamperuma 1971); the same remark can be made for proteinoids produced by condensation of the N-carboxy amino acid anhydrides (Hayakawa et al.1967). The composition of the material polymerized from amino acyl adenylates was found to reflect the starting mixtures (Banda and Ponnamperuma 1971). This result is further evidence of some lack of randomness in polypeptides which are synthesised abiotically (Fox and Harada 1960b; Fox and Nakashima 1967).

Of course the most obvious source of non randomness in the absence of a template is neighbour interaction and different activiti between amino acids . This possibility was confirmed by Steinman and Cole (1967) who measured the frequencies of formation of dipeptides in mixtures of amino acids, using glycine-glycine as a standard (table II-3). From a comparison of their results with those calculated from known protein sequences (also table II-3), Calvin (1975) argues that "the original polypeptide was similar to the ones we have today". Analogous results were obtained by Paecht-Horowitz (1974) with adenylates (table II-4).

TABLE II-3

Dipeptide yields, as determined experimentally and as calculat
from known protein sequences. After Steinman and Cole (1967).

	Frequencies (relative to Gly-Gly	
Dipeptide	Experimental	Calcul
Gly-Gly	1.0	1.0
Gly-Ala	0.8	0.7
Ala-Gly	0.8	0.6
Ala-Ala	0.7	0.6
Gly-Val	0.5	0.2
Val-Gly	0.5	0.3
Gly-Leu	0.5	0.3
Leu-Gly	0.5	0.2
Gly-Ile	0.3	0.1
Ile-Gly	0.3	0.1
Gly-Phe	0.1	0.1
Phe-Gly	0.1	0.1

TABLE II-4

Relative yields of bonds in the copolymerization reactions of adenylates of pairs of amino acids. After Paecht-Horowitz (197

Interacting substances	Bonds	Relative yields of bonds (%)
alanine-adenylate glycine-adenylate	Al-Al	40
	Gly-Gly	32
	Al-Gly	15
	Gly-Al	13
alanine-adenylate valine-adenylate	Al-Al	23
	Val-Val	52
	Al-Val	12
	Val-Al	13
alanine-adenylate aspartyl-adenylate	Al-Al	47
	Asp-Asp	49
	Al-Asp	2
	Asp-Al	2
alanine-adenylate serine-adenylate	Al-Al	37
	Ser-Ser	37
	Al-Ser	12
	Ser-Al	14
aspartyl-adenylate glycine-adenylate	Asp-Asp	55
	Gly-Gly	21
	Asp-Gly	9
	Gly-Asp	15
aspartyl-adenylate serine-adenylate	Asp-Asp	59
	Ser-Ser	22
	Asp-Ser	10
	Ser-Asp	9
aspartyl-adenylate histidyl-adenylate	Asp-Asp	36
	Hist-Hist	44
	Asp-Hist	8
	Hist-Asp	12

Later I will explore the implications of these findings for the origin of the genetic code.

Properties of polypeptides produced in the absence of a template read-out system (proteinoids)

All types of proteinoids produced exhibit remarkable simi larities with natural proteins. Whether or not a clear prebio pathway to their production has been established to everybody' satisfaction, they represent a plausible intermediate stage in the evolution of matter, and as such they can be used as a new starting point in the study of prebiotic systems.

In spite of the limited heterogeneity reported above, the eighteen most common amino acids have been incorporated both in thermal proteinoids (Fox et al. 1963, Fox 1969) and in thos produced by the condensation of adenylates (Krampitz and Fox 1969). The relative proportions of amino acids are found to fo low rather closely those of an average protein (table II-5).

Thermal proteinoids have exhibited catalytic activity: decomposition of glucose (Fox and Krampitz 1964), hydrolysis of p-nitrophenyl acetate (Fox et al. 1964; Rohlfing and Fox 1967, Shook and Rohlfing 1972), hormone-like stimulation of melanocyt (Fox and Wang 1968), catalysis of several simple reactions (Roh fing and Fox 1969, Dose 1971). With the incorporation of zinc, proteinoids have also shown activity for the hydrolysis of ATP

TABLE II-5

Composition of hydrolysate of proteinoid from amino acid adenylates alone compared with an average protein (calculate without ammonia). After Krampitz and Fox (1969).

Amino Acid	Composition of polymer (mole %)	Ratios of amino in average prote (mole %)
Lysine	6.5	5.9
Histidine	2.4	1.8
Arginine	4.2	4.9
Aspartic acid	10.3	9.7
Threonine	4.9	4.8
Serine	4.2	6.0
Glutamic acid	9.7	12.7
Proline	5.1	6.2
Glycine	11.1	12.6
Alanine	14.3	9.6
Valine	7.3	5.9
Methionine	0.7	1.8
Isoleucine	4.5	6.0
Leucine	9.6	6.0
Tyrosine	0.1	2.3
Phenylalanine	4.5	3.7

* Average protein composition from Vegotsky, A., and S.W. Fox in Comparative Biochemistry, ed. M. Florkin and H.L. Maxon (New York: Academic Press, 1962), vol. 4, p. 185.

(Ponnamperuma and Young 1964). The activity has been demonstr
to depend on specific amino acid residues (Fox and Wang 1968)
in particular on the presence of histidine and imide linkages
aspartic acid (Rohlfing and Fox 1967); it is also maintained (c
even increased) after storing in dry state (Rohlfing 1970).

Table II-6 lists the properties which can be associated both with thermal proteinoids and proteins.

Of particular interest is the tendency of proteinoids to associate and form "microspheres", and to exhibit new properti as a result of being thus organized. I will report on micro-spheres and on the interactions of proteinoids with nucleic ac in sections to follow.

Are thermal proteinoids the actual ancestors of today's proteins? The experiments being performed prove that many wor are still looking for alternatives to thermal condensation. Very recently Andini et al. (1975) have shown by NMR that, con-trary to the data based on chemical degradation, only β-peptid linkages are present in thermally synthesized poly-aspartic acid. This in their words, casts "serious doubt on the role of thermal condensation in prebiotic synthesis". From the point of view of this work, however, I need not enter the controversy.

TABLE II-6

Properties common to thermal proteinoids and contemporary proteins and/or enzymes. Adapted from Fox (1969, 1974a). A related bibliography can be found in the second publication.

Qualitative composition

Quantitative composition (except serine, threonine)

Range of molecular weights (4,000-10,000)

Color tests

Inclusion of nonamino acid groups

Solubilities

Salting-in and salting-out properties

Precipitability by protein reagents

Hypochromicity

Infrared absorption maxima

Recoverability of amino acids on mineral acid hydrolysis

Susceptibility to proteolytic enzymes

Many catalytic activities

Nutritive quality

Tendency to assemble into microparticulate systems

Limited heterogeneity

Inactivability by heating in aqueous solution

Hormonal activity (melanocyte stimulation)

Binding of polynucleotides (by basic proteinoids)

pH-Activity curves

Michaelis Menton Kinetics

Inhibition and reversal

Specificities

II.3.2 Synthesis of Oligonucleotides in the Absence of Enzymes

Synthesis without template

Only oligonucleotides have been produced under prebiotic conditions. Internucleotide phosphodiester bonds were observed by Schwartz and Fox (1964, 1967) and Schwartz et al. (1964) whe cytidine 2'(3')-phosphate was heated in the presence of poly- phosphoric acid; and there is evidence that dinucleoside phosph were formed in an experiment on nucleotide synthesis I mentione earlier (Ponnamperuma and Mack 1965). The same experimenters s gested that clays might aid the polymerization process. Morave et al. (1968a,b) reported yields of di- and trinucleotides in t thermal condensation of uridine. They observed both 2'-5' and 3'-5' bonds with some preference for the natural 3'-5'. Tapier and Nagyvary (1971) heated the triethylammonium salt of cytidine 2', 3'-cyclic phosphate for 48 hrs. at 138°C to obtain oligomer of cytidylic acid up to hexamer. Their dimers showed about equal amounts of 2'-5' and 3'-5' bonds, but the trinucleotides had a preference for the 2'-5' (66%). Similarly, in trinucleo- tides formed in a uridine, urea, ammonium dihydrogen phosphate system, Osterberg et al. (1973) found 60% 2'-5' bonds, while th same percentage applied in favour of the natural linkage in di- nucleotides. The presence of "unnatural" 2'-5' (and 5'-5') phos phodiester bonds is even enhanced in template directed synthesis (see below), but "this need not concern us greatly" comments Lehninger (1970, p. 779), "the primordal nucleic acids might well have had 2',5' linkages". Citing evidence from Kondo et al (1970), Oro and Stephen-Sherwood (1974) concur:

> It can be anticipated that higher oligomers would possess random bonding with both 3'→5' and 2'→5' bonds present in the chain. These oligomers would not necessarily be ineffective in functioning as a primitive template, since in solution both the 3'→5' and the 2'→5' dimers of adenyladenine stack with an anti-anti right-handed conformation typical of RNA polymers.... The selection of the natural linkage may have occurred at some later stage in the development of a self-replicating system.

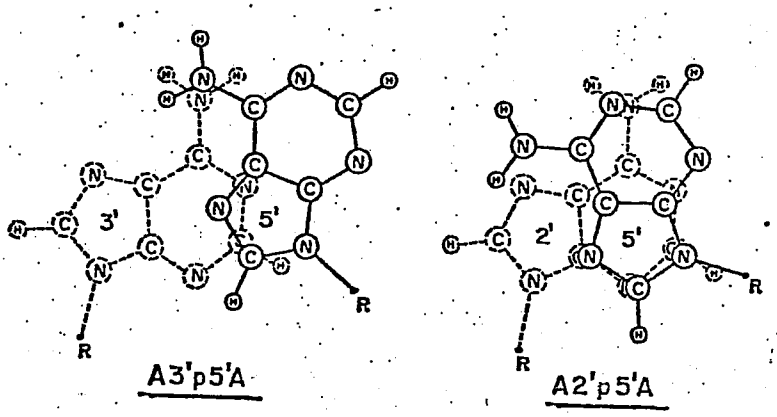
(see Fig. II-2).

Oró and Stephen-Sherwood (1974) feel that a urea-ammoniu chloride-phosphate system, while satisfactory for the synthes of oligoribonucleotides, is not applicable to the forma- tion of oligodeoxyribonucleic acids. One possible approach makes use of a deoxynucleoside 3',5'-cyclic phosphate as an i mediate in oligonucleotide synthesis (Pong and Tso, 1971). I et al. (1971) reported the condensation of nucleotides by hea TMP in an aqueous solution of cyanimide and, in the presence montmorillomite, they detected oligodeoxyribonucleotides up t 5 units long (interestingly though, fewer di- and trinucleoti were found when the clay was used). Oro and Stephen-Sherwood (1974) consider it likely that thymidine 3',5'-cyclic phospha was an intermediate "although a more direct cyanimide dehydra reaction cannot be eliminated." They can also envision the u of preactivated mononucleotides and review the evidence in su port of this approach.

The relatively modest results obtained in this area are viewed by many researchers (for instance, Fox and Dose 1972) a an indication that proteins and catalytic activity had to prec the formation of nucleic acids. If this is the case, then

FIGURE II-2

Schematic presentation of the front view of the conformationa models for A_3,p_5,A, A_2,p_5,A (bases drawn by the dotted line a located at the bottom). The axis is advancing upward from th plane of the paper and is rotating counterclockwise simultane ously. After Kondo et al. (1970).



proteinoids and microspheres (above and sct. II.4.1) are obv
ous new research tools; and some evidence has already been
presented (Fox et al. 1974) to show that proteinoids and mic
spheres can indeed aid in the polymerization of nucleotides
table II-7).

Template directed synthesis

Experiments in this area reveal a few interesting facts

a) Purine-pyrimidine complementarity guides the polymeriza
process even in the absence of enzymes.

b) Triple stranded helices (2 polymer strands aiding in th
formation of a third strand from monomers) are favoured
thermodynamically (Hoffman and Pörsche 1973a,b).

c) Complementary helices are formed only by purine monomer
on pyrimidine templates (Orgel, 1968; Orgel and Sulston,

d) A "stacking interaction" among the monomers favours the
condensation on polymer templates (Renz et al. 1971; Pör
and Eggers, 1972).

Howard et al. (1964) observed the formation of helices
poly-C and guanosine mononucleotides. They found the 2:1 sto
ometry typical of triple helices as well as, for well-defined
conditions, a 1:1 stochiometry for double strands. Some of t
same experimenters (Howard et al. 1966) were only able to det
triple strands for poly-U to adenosine, and a threshold effec
in the concentration of monomers (Felsenfeld and Miles 1967),
which is probably due to the stacking interaction mentioned a
1:1 stochiometry for binding of adenosine to poly-U was inste
found by Huang and Tso (1966) at 20°C.

TABLE II-7

Yields of adenine dinucleotide and trinucleotide in various systems with ATP (at 37°, 2 days). After Fox et al. (1974)

Products	Reactants			
	ATP in aqueous solution	ATP with basic proteinoid	ATP with acidic proteinoid microspheres	ATP acid prot micr
Adenine, adenosine, cAMP, AMP, ADP	12.6%	14.7%	14.8%	13
ATP recovered	86.8	82.9	82.9	84
Oligo A eluting beyond ATP	0.7	2.2	2.2	2
Trinucleotide	0.0	0.0	0.2	0
Dinucleotide				

Under conditions similar to those favoring the formation triple helices other workers observed the synthesis of oligo- adenylic acid from adenylic acid (Sulston et al. 1968a,b) and similarly of oligonucleotides of deoxyadenylate (Schneider-Be loehr et al. 1968). They found mainly 2'-5' and 5'-5' linkage when the complementarity rule was respected, i.e. between ade sine residues, whereas poly-U had no effect on the binding of adenylic acid with guanosine, cytidine or uridine (Sulston et 1968b). Experiments with a poly-C template showed similar ef on guanosine and none on the other nucleosides (Sulston et al 1969). The formation of internucleotide bonds was made possi by the presence of a water soluble carbodiimide as condensing agent. The same group did not consider this a plausible prel agent and proposed instead a prebiotic model which made use o imidazoles. They observed the formation of internucleotide b with adenosine-5'-phosphoimidazolide and a poly-U template (Weimann et al. 1968).

However, according to Hoffman (1975), only Porsche et al (1973) have clearly demonstrated that monomers can form a double helical complex with a complementary template.

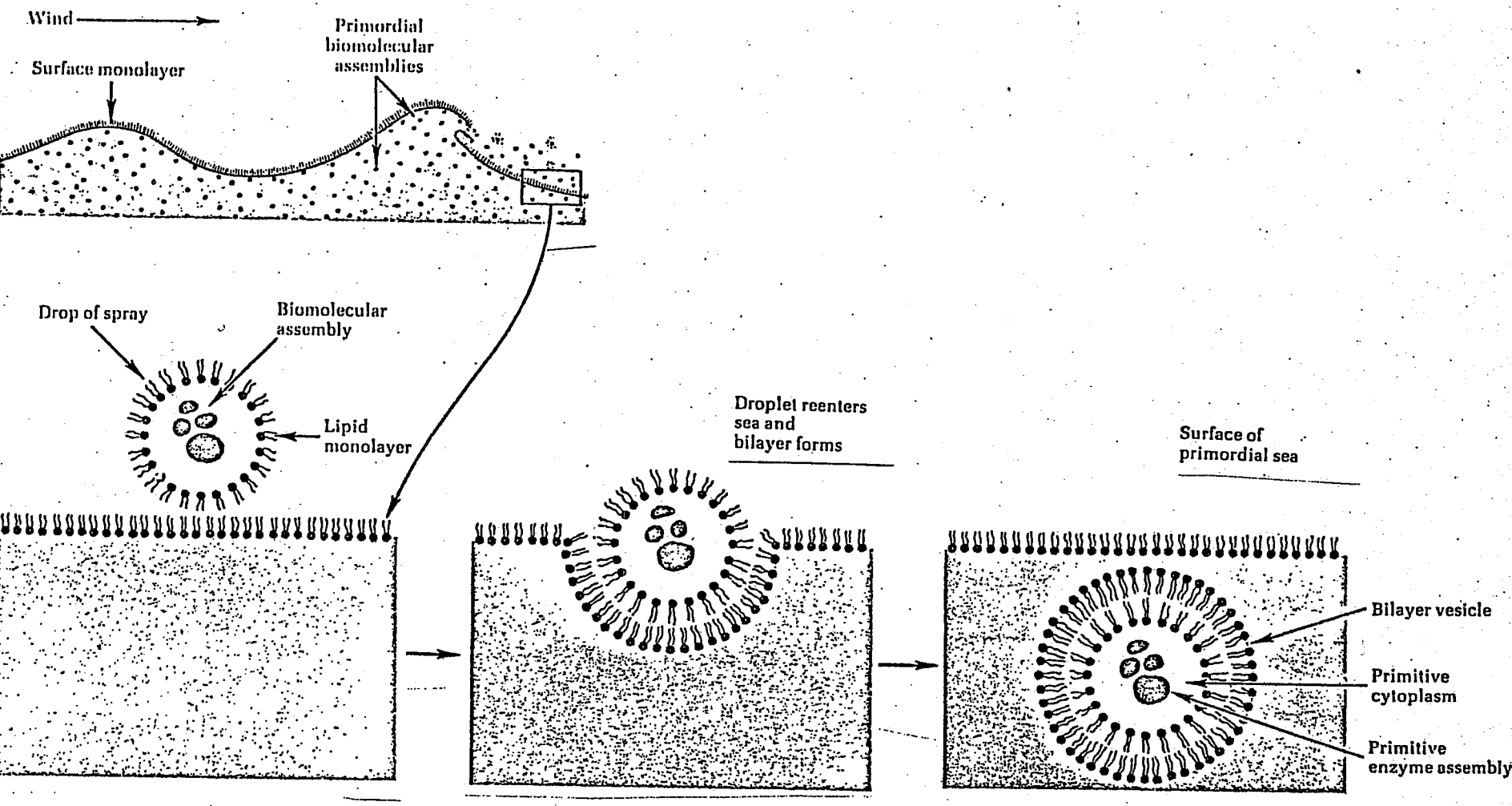
From evidence on enzymatic synthesis Oro and Stephen- Sherwood (1974) argue that oligonucleotides 9-12 units in len could serve as primers for their own replication, and propose a number of ways in which further elongation might be achieve Unfortunately one can only speculate at this point.

A triple-stranded polynucleotide helix containing only purine bases

An apparent exception to the pyrimidine-purine pairing in helices has been found: two polynucleotide strands of hypoxanthine can form a stable helix with one polynucleotide strand of adenosine residues (Arnott and Bond, 1973). The experiment reported was not based on a model of prebiotic synthesis, and perhaps the result is not surprising if one considers that A-I pairing has long been known for codon-anticodon interactions (Crick 1966); however, the possibility of purine-pyrimidine complementarity evolving from such a model without major discontinuities, as suggested by Arnott and Bond, is an attractive possibility. Orgel and Crick (Crick, 1968) had already propos such a scheme, having inferred from experiments in prebiotic synthesis that adenosine phosphate and its deaminated derivati inosine phosphate were probably the most abundant prebiotic nucleotides. Clearly more experiments are needed before the i portance of purine-purine complementarity for prebiotic system can be definitely established.

FIGURE II-3

Formation of lipid bilayer membranes. After Lehninger (1970, p. 785)



Evreinova et al. (1974) state:

> A coacervate system can be obtained from synthetic polymer molecules, e.g., from polyadenylic acid and polylysine (according to Oparin and Serebrovskaya), and also from biopolymer molecules: proteins, nucleic acids, carbohydrates, and other compounds (2). To date over 300 chemically different coacervate systems are known. Enzymic reactions have been carried out in many of these

---

2. Evreinova, T.N., "Concentration of Substances and Action of Enzymes in the Coacervates", "Nauka", Moscow (in Russian), 1966.

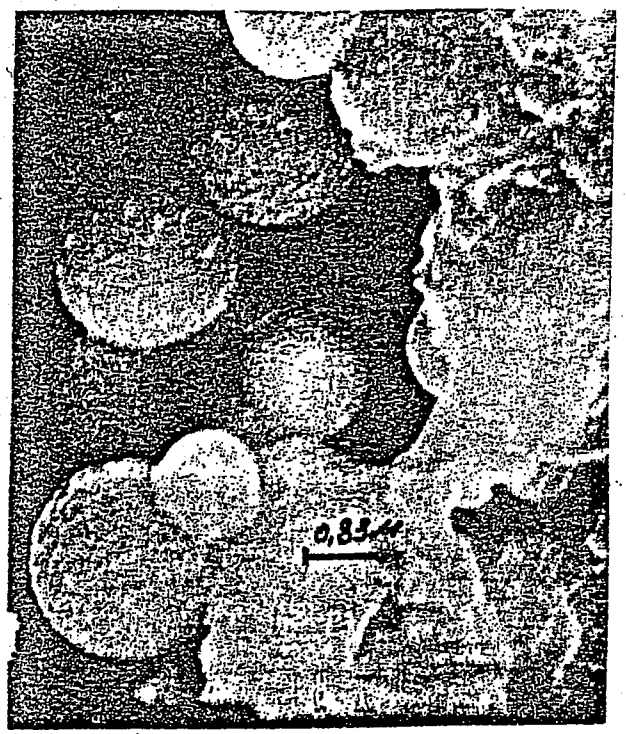
Fox (1965) pointed out that a typical coacervate droplet "mad from gelatin and gum arabic cannot withstand gentle centrifug or some concentrations of salts without breaking down into tw liquids". He further objected to the reliance of such a mode proteins, like gelatin of recent origin. However drops that lasted for years have now been reported (Evreinova et al. 197 in fact according to those authors' statement, quoted above, Fox's "microspheres" (following section) can be viewed as a pa cular coacervate system. Other Russian sources follow the sa convention, for instance Novak (1974). A clarifying statement by Fox will be quoted below.

Typical coacervate drops (Fig. II-4,a,b) vary in diameter between 0.5 and 640 microns (Evreinova et al. 1974), can favou enzymatic processes in their interior (Oparin, 1966; Lehninger 1970, p. 782) and have a tendency to form colonies when they c in contact (Fig. II-5). Their sizes, masses and concentration dry material are compared with those of living systems in Tabl II-8.

FIGURE II-4

Coacervate drops in the scanning electron microscope. Afte
Evreinova et al. (1974).

a) Coacervate system: polyphenoloxidase-histone-gum arabic-
quinones



b) Coacervate system: peroxidase-histone-gum arabic-
purpurogallin.

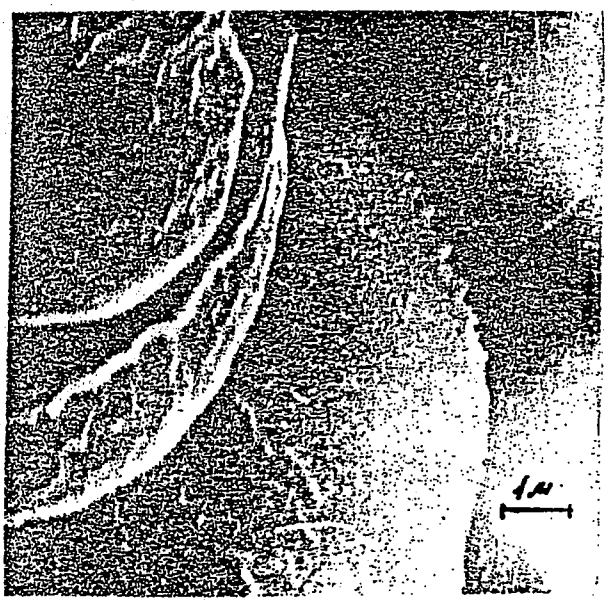


FIGURE II-5

A colony of coacervate drops in the scanning electron microscop
Coacervate system: polyphenoloxidase-histone-DNA-guinones.
After Evreinova et al. (1974).

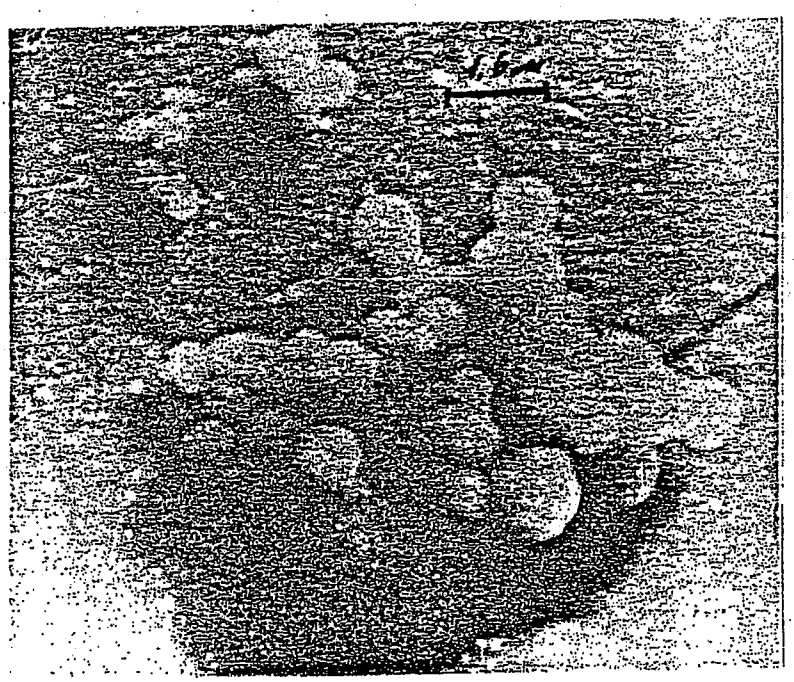


TABLE II-8

Weight and concentration of dry mass in various objects. Afte
Evreinova *et al.* (1974).

Objects	Diameter (cm)	Volume (cm^3)	Weight (g)	Conce tion
Bacteria	2.5×10^{-4}	8.5×10^{-12}	2.5×10^{-12}	30-2
Mammalian cells	2.5×10^{-3}	8.5×10^{-9}	2.1×10^{-12}	25-1
Amoeba	1.0×10^{-2}	5.2×10^{-7}	7.8×10^{-8}	15-1
Coacervate drops	2.42×10^{-4} to 1.02×10^{-2}	7.4×10^{-12} to 5.3×10^{-7}	2.5×10^{-12} to 3.5×10^{-8}	34-7

Proteinoid microspheres

> The manner in which the proteinoid microspheres aris is yet simpler than that by which the polymer emerges. Water or various aqueous solutions are added to the hot polymer mixture, the hot clear solution is decanted and, following a few minutes of cooling, vast numbers of individual microspheres are seen to separate.

A typical result of the procedure just quoted (Fox, 1965) can seen in Fig. II-6. More details and different procedures can found in several articles (for instance Fox, 1964a,b; Fox and Yuyama, 1964; Miquel et al., 1974).

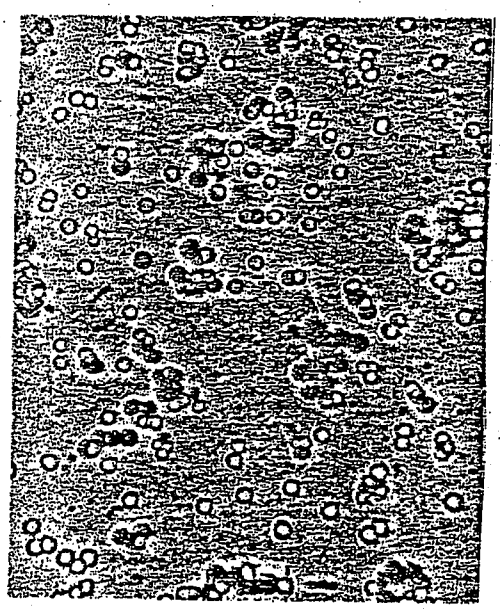
Having been asked to compare and contrast coacervate drop and microspheres during an interacademy cultural lecture exchange between Oparin and himself in 1969, Fox (1974b) re

> The contest of those questions indicated that coacervate droplets from gelatin and gum arabic were usually meant on the one hand, whereas microspheres assembled from proteinoid alone were meant on the other. Difficulties stem from the fact that either coacervate droplets or proteinoid microspheres are produced in many types. Coacervate droplets so designated are complexes of two or more polymers of markedly different composition. The simplest proteinoid microsphere consists mainly of a family of compositionally closely related copolyamino acids, rather than of two colloidal types of markedly different sources and polarities. It is also true that coacervate droplets of, for example, polyvinyl sulfonic acid... alone have been reported, but this type is not relevant to a question of associations of functions related to vital properties, associations such as are found in copolyamino acid preparations. A fundamental difference between the typical coacervate droplets of the Oparin school and the typical proteinoid microspheres is that the former are produced from polymers obtained from contemporary organisms, whereas microspheres are aggregates of proteinoid arising in turn from monomeric amino acids under one set of geologically relevant conditions...

Microspheres were found to possess a number of very remarkable properties and morphologies which in many cases were "inadverte

FIGURE II-6

Proteinoid microspheres prepared by allowing a hot solution o proteinoid to cool. These units are approximately 2μ in diameter. After Fox (1965).



obtained" (Fox, 1965). For instance multiple layer surfaces (apparently 2 layers in Figs. II-7 and II-8, and several in Fig. II-9), ability to propagate by cleavage (Fig. II-10) and through budding (Fox et al., 1967; Fig. II-11). Brownian moti has been observed in microspheres (Fox and Yuyama, 1964), together with internal structural constraints.

Very important, of course, is their catalytic ability, as shown in many experiments (Fox, 1974a); for instance those of Fox and Krampitz (1964), Ryan and Fox (1973), and Fox et a (1974). In this last experiment (which I briefly referred to in the section on the synthesis of oligonucleotides) ATP was added to microspheres of basic and acidic proteinoids suspend in 20 mM $MgCl_2$ solution, to yield adenine dinucleotides and trinucleotides. The experimenters point out that "a definite peak for the trinucleotide has been found only in the microsphere systems, under conditions that do not yield adenine trinucleotide in the absence of such particles". The reader may wish to see Table II-7 again.

Structure and properties of microspheres have been chang somewhat by new experimental conditions such as the inclusion of nucleic acids (Wachneldt and Fox, 1968).

To allow for all possibilities Fox (1969) has begun to u the name "microparticulate units" or"microparticles". With t name the properties of "old" and "new" microspheres are liste together in Table II-9.

Of the most recently synthesized microparticles some can viewed as models of primitive ribosomes, which I will discuss

FIGURE II-7

Optical micrograph of microspheres. In the source it is st
"What looks like concentric shells, or a kind of double lay
may also be seen." The filaments are associations of small
microspheres. After Fox (1965).

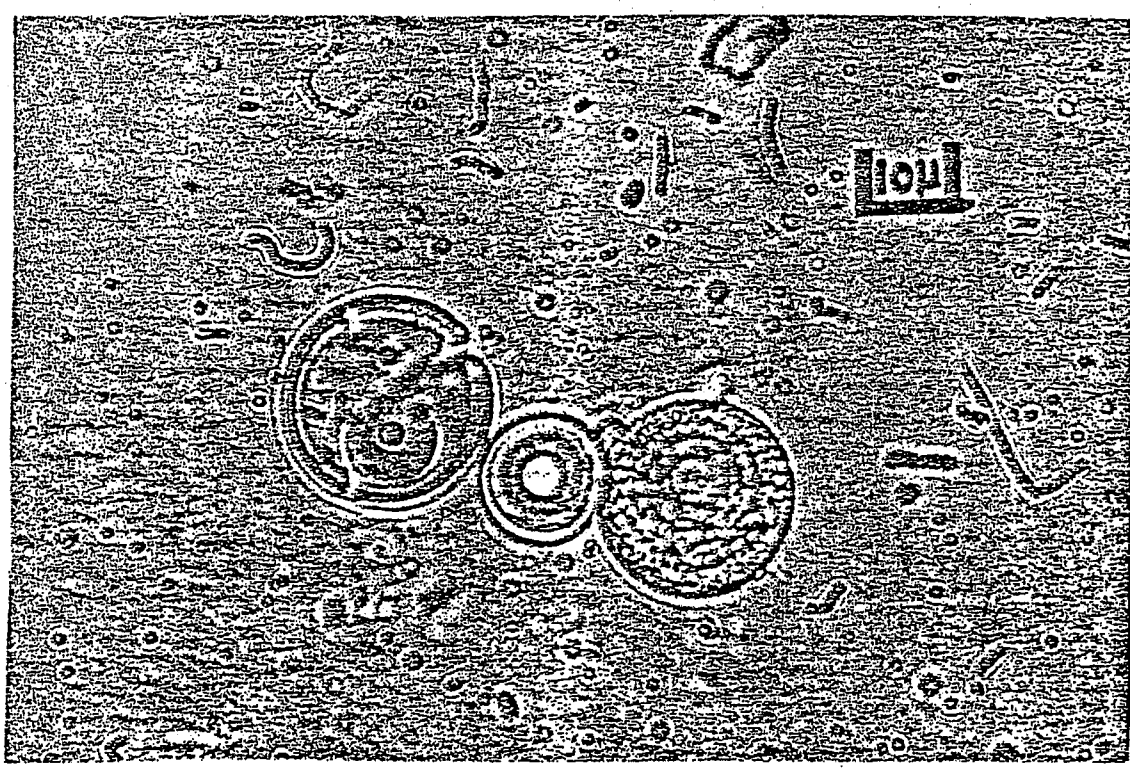


FIGURE II-8

Electron micrograph of microspheres. Double layers are clear visible. After Fox (1965).

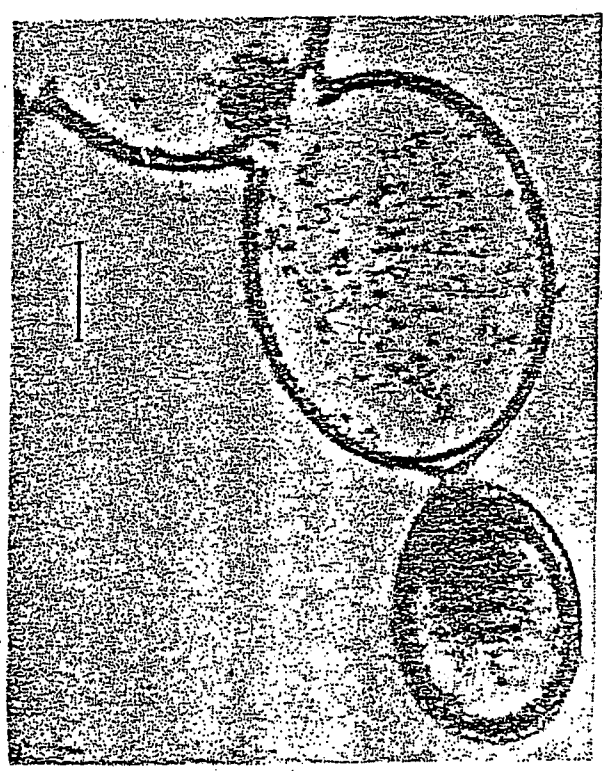


FIGURE II-9

An electron micrograph by Professor Walther Stoeckenius. The proteinoid microsphere is 1 micron in diameter. Concent boundary layers are visible. After Fox (1975).

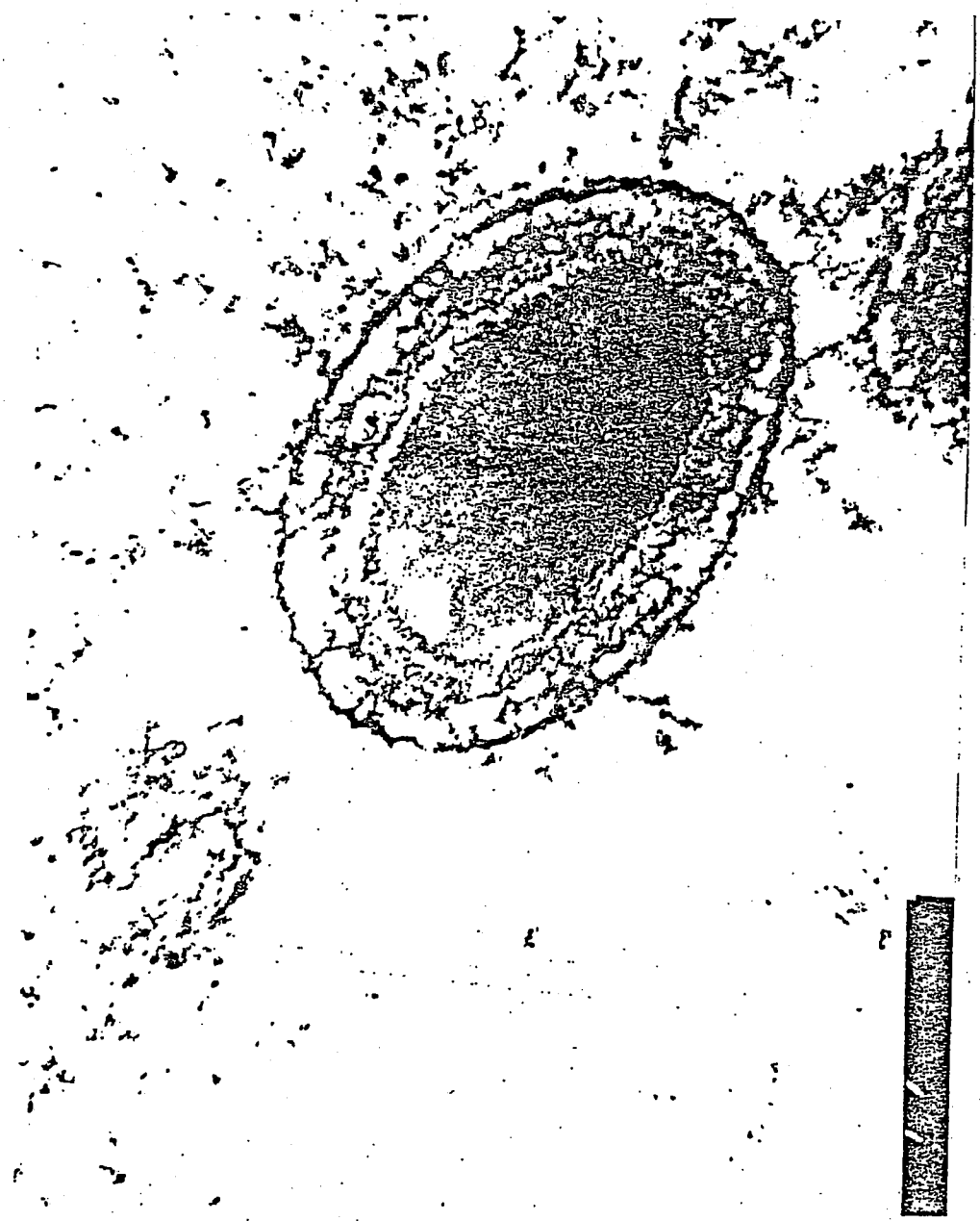


FIGURE II-10

Microspheres simulating cells in cleavage. The author presen
evidence that they are not dividing bacteria and that indeed
the process is cleavage rather than fusion. After Fox (1965)

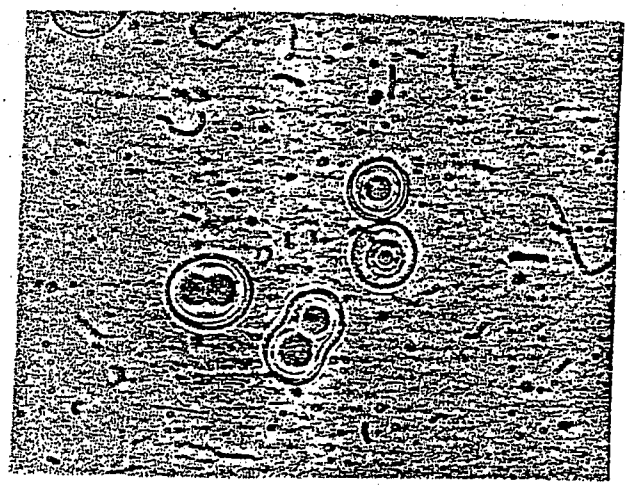
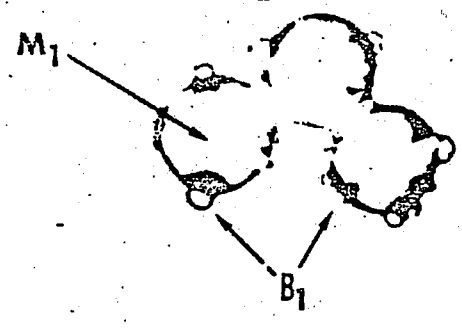


FIGURE II-11

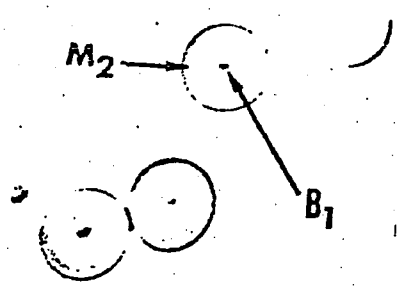
Time sequence a) through d) of proliferation through budding. After Fox (1975).

a) proteinoid microspheres (M_1) with buds (B_1);



b) buds (B_1) have separated;

c) second generation microspheres (M_2) are grown by accretion around separate buds (B_1);



d) a second generation bud (B_2) on a second generation microsphere (M_2).

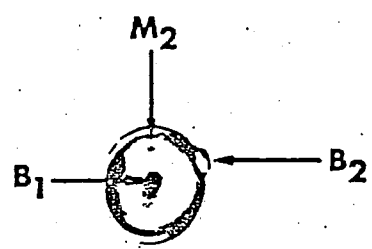


TABLE II-9

Properties of proteinoid microparticles:

a) microspheres described by 1965
b) units made of proteinoids since 1965

After Fox (1969).

a) Stability (to standing, centrifugation, sectioning

Microscopic size

Variability in shape

Uniformity of size

Numerousness

Stainability

Producibility as gram-positive or gram-negative

Osmotic type of property in atonic solutions

Structured boundary

Ultrastructure (electron microscope)

Selective passage of molecules through boundary

Catalytic activity

Patterns of association

Budding and fission

b) Motility

A number of catalytic activities

Growth by accretion

Ability to propagate through budding and growth by accretion

Binding of polynucleotides in various ways

in section II.4.2, others, like those synthesized by Rohlfing (1975) are more resemblant of coacervate droplets.

Extraterrestrial and fossil evidence

Claus and Nagy (1961), in examining some carbonaceous chondrites, found "organized elements" which they interpreted as "possible remnants of organisms" probably indigenous to the meteorite. In line with a suggestion of Morrison (1962), and impressed by the striking similarities between these remnants and his microspheres, Fox (1965) suggested that:

> .... the 'formed elements' were at no time alive, rather that they were natural physico-chemical experiments which terminated before life emerged.... [the similarities between the organized elements and microspheres] are consistent with the inference that the laboratory experiments are closely akin to natural experiments [italics mine].

Terrestrial fossil remnants have also been found which resembl coacervate droplets and microspheres (for instance, Mueller, 1972).

II.4.2 Interactions Between Amino Acids and Nucleotides

As the existence of a genetic code became clear, it was natural to ask whether the relationship between codons and amino acids was stereochemical. Soon evidence was found that

this was not the case, and an RNA adapter was needed (Chapeville et al. 1962; Chapeville 1962; Chapeville et al. 1963); nevertheless Pelc and Welton (1966) proposed a number of dire fits and suggested that the results of Chapeville et al. coul be explained by "overlap of fits" (Welton and Pelc 1966), but Crick (1967a) pointed out that their models had been built backwards, thus really showing fits for the wrong amino acids Dunnill (1966) proposed instead that the amino acids could fit stereochemically to the anticodons, and Melcher (1970, 1974) constructed models to support the idea (fig. II-12). However, a number of studies has shown that modification of t anticodon loop of the tRNA does not prevent attachment to the correct amino acid, i.e. the synthetase recognizes other parts of the tRNA or possibly its overall shape (Gauss et al., 1971) It is likely that the techniques will soon be developed to clarify these and other needed details of the workings of the genetic machinery.

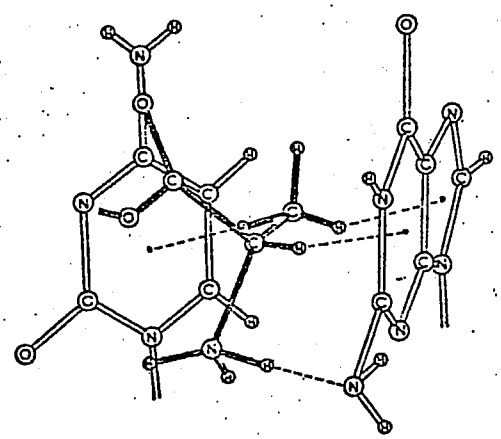
The determination of the 3-dimensional structure of the phe-tRNA of yeast (fig. II-13) is a good example of the progre that is being made. Yet there is still no indication that, at present, either the codon or the anticodon site help recognize the amino acid.

Assuming that the tRNA functions as adaptor, one can still ask whether its adaptor function is itself a product of

FIGURE II-12

Examples of postulated stereochemical fits between amino aci
and anticodons. After Melcher (1974).

a) Alanine, anticodon: C-G-Pu,Py



b) Serine, anticodon: A-G-Pu,Py

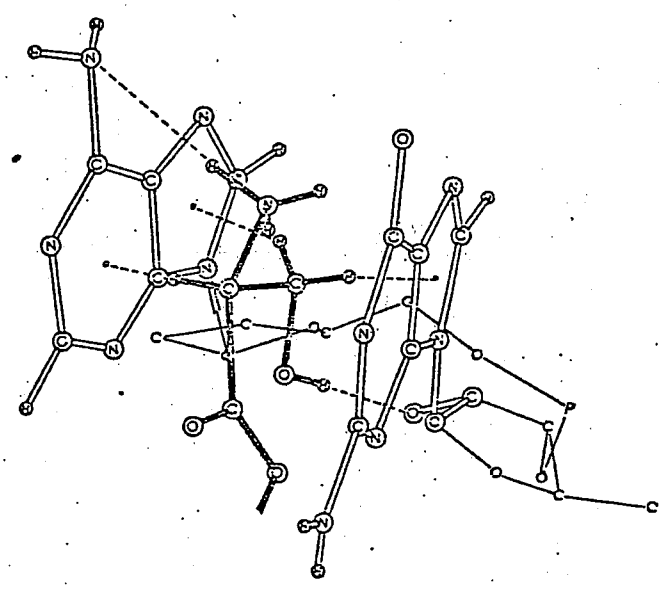
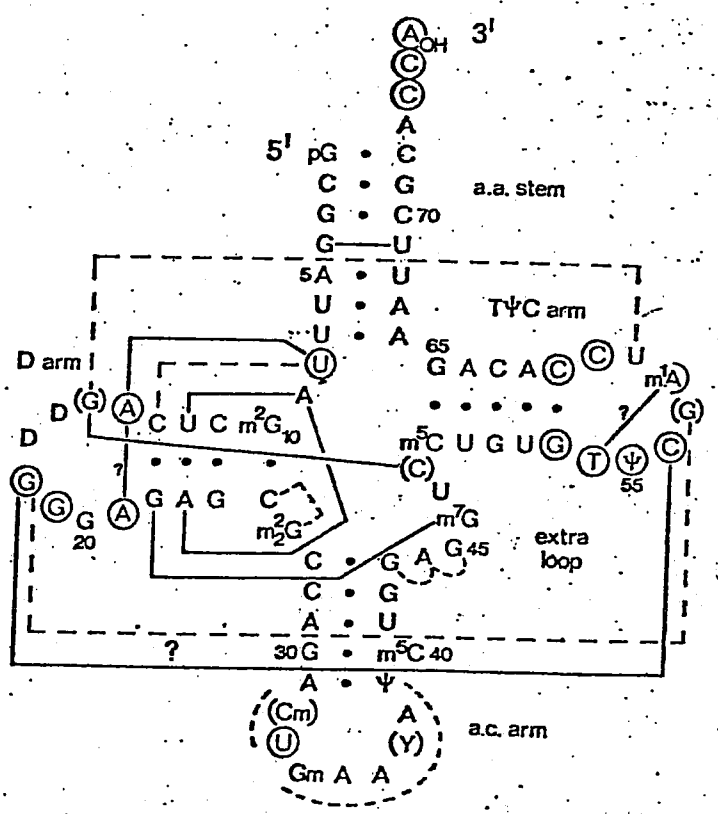


FIGURE II-13a,b

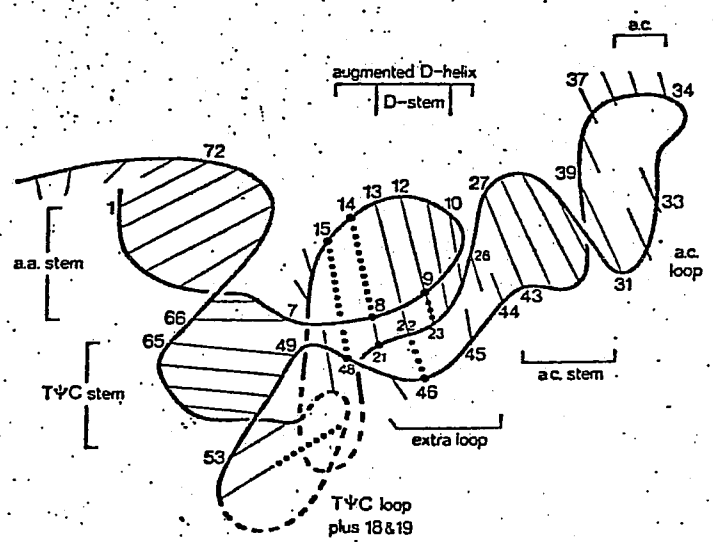
Structure of Yeast tRNA^{Phe}

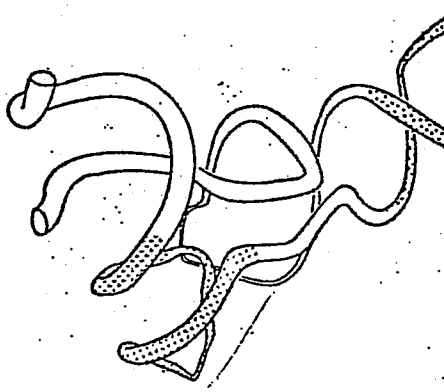
a) The sequence is arranged in the clover leaf formula. Cir
indicate bases which remain invariant in all tRNA sequer
brackets indicate those which are always either purines
pyrimidines (semi-invariants). Solid lines join bases w
are paired in the tertiary structure, and dashed lines j
those which are stacked on each other. After Robertus e
(1974).



b) Schematic diagram of the tertiary structure (left). Th
tinuous line represents the ribose-phosphate backbone.
light lines represent double helical stems, shorter lin
represent non-paired bases. The dashed and the dotted
represent, respectively, ambiguities in the tertiary stru
and base pairs which are normally not indicated in the
leaf formula.

The three-dimensional structure is made more eviden
comparison with a perspective diagram of the MIT model
which however differs in some respects. Adapted from R
et al. (1974) and Kim et al. (1973).





evolution or was necessary from the very beginning (i.e., th correct adaptors get together by "accident").

The quotations I placed at the beginning of this disser illustrate the ongoing controversy between these seemingly incompatible positions. Later I will try to show to what exte they are really incompatible. In any case, as suggested by Crick (1968), interactions between codons or anticodons and amino acids "may have been important in the past at least fo a few amino acids." Is there evidence of such interactions?

Selective binding of mono- and polynucleotides

Early experiments with histone-like structures showed that polylysine bound preferentially to DNA samples which we rich in A-T pairs (Spitnik et al. 1955, Ohba 1966). Polyarginine showed instead a "marked preference" for poly G-C (Leng and Felsenfeld 1966). Sober et al. (1966) studied the effect of nuclease on polylysine-ribonucleic acid complexes and found that RNA with high G-C content was more "protected" by polylysine; they pointed out however that the protection specificity they observed did "not necessarily reflect binding specificity". They gave alternative explanations and added that even if protection specificity did reflect binding speci ficity, it did "not appear to be high enough to be of biological consequence"; but they also felt that this was partly due to their particular experimental conditions and that "in

biological systems, proteins with separated short runs o basic amino acids might possess a great deal of informat for nucleic acid sequence specificity".

Latt and Sober (1967a) also found that polylysine b more strongly to poly I-C than to poly A-U, but observed that true specificities could only be detected if the le of the polylysine chain was kept constant (binding was l dependent) and that "in the presence of different cation [the specificity of the interaction was] markedly influe by the nature of the cation" (Latt and Sober 1967b). In fact they emphasized: "The oligolysine binding preferen seen in NaCl solutions appears to be largely a consequen of the reverse specificity of the Na^+". Discrepancies i past results could probably be ascribed to these factors.

Wagner and Arav (1968) studied the interaction of mo nucleotides with polylysine and polyarginine and found th differences in binding behaviour could be detected betwee purine and pyrimidine nucleotides, and between individual purine nucleotides. Guanylic acid bound most strongly in every case. Their conclusion was that

> ... there is some specificity in the interaction tween a nucleotide and a basic polypeptide. This sp cificity may be due to nucleotide stacking and/or to nonelectrostatic affinities between a nucleotide and its binding site superimposed on the major electro- static interaction.

Woese (1968) emphasized instead the role of stacking actions. He pointed out that non-specific electrostatic between the phosphate groups of the nucleotides and the si

chains of the polyamino acids could be responsible for pos
ing the nucleotide monomers, while their differences in st
interactions (G > A > C > U from strongest to weakest) w
naturally appear as different percentages of bound monomer
His experiments and those of Lacey and Pruitt (1969) confi
the effect predicted above. But the possibility of an add
tional, more specific, nonelectrostatic effect had not bee
ruled out.

Waehneldt and Fox (1968) prepared particles from lysi
rich proteinoids and organismic RNA or thermally synthesis
oligocytidylic acid. In similar experiments, Yuky and Fox
found that arginine-rich (lysine-free) and lysine-rich (ar
free) proteinoids reacted more readily respectively with p
and with pyrimidine homopolynucleotides (table II-10).

These and other studies led Saxinger and Ponnamperuma
(1971) to state that

> a variety of interactions can occur, depending upon composition, conformation, state of polymerisation, and environment of the reacting species.

They proposed that a systematic study of these interaction
to start at the simplest possible level (both species monom
although the results could not be expected to "be spectacul
from a biological viewpoint". For this purpose Saxinger et
(1971) immobilized amino acids on a chromatographic support
measured the binding of ribonucleoside-5-monophosphates. T

TABLE II-10

Binding of polynucleotides to proteinoids. As deduced by formation of microparticles, polyribonucleotides bind pre tially with either lysine-rich or arginine-rich proteinoi After Yuki and Fox (1969).

Polyribonucleotide	Lysine-rich (arginine-free) Proteinoid		Arginine-rich (lysine-free) Proteinoid	
	Turbidity	Number of Spherules	Turbidity	Number Spheru
Poly C	0.253	1.0×10^5/ml	0.002	
Poly U	0.050	7.5×10^2/ml	0.058	1.3 x
Poly A	0.001	-[a]	0.060	1.4 x
Poly G	0.003	-	0.218	4.2 x
Poly I	0.003	-	0.248	1.0 x

[a]-Signifies < 1 spherule/ml.

concluded that

> Even on the monomer level, interactions between amino acids and nucleotides are sufficiently strong and of such nature that a rudimentary "preferential" scheme can be seen.

Similar experiments, performed with oligonucleotides showed a "marked dependence on the composition, size, and order of bases within a given oligo nucleotide" (Saxinger and Ponnamperuma 1974, table II-11).

Selective binding of amino acids

Woese et al. (1966) subjected amino acids to paper chromatography with solutions of pyridine to find evidence for organic base-amino acid interactions. They defined as "polar requirement" a function of the R_f (fraction of distance traveled by the amino acid) and at the mole fraction of water in the pyridine solution. There appeared "to be a striking correspondence between a polar requirement ordering and a codon ordering of the amino acids". And Woese and colleagues continued:

> Since the former ordering is based upon pyridine-amino acid interactions, we feel the conclusion is essentially unavoidable that the codon assignments manifest an underlying codon-amino acid pairing.

Harpold and Calvin (1968, 1973) attached either adenine or cytosine to polystyrene, a synthetic polymer, and measured the efficiency with which either phenylalanine or glycine were bound. Glycine was more reactive than phenylalanine and both showed preference for adenine (table II-12).

TABLE II-11

The binding of oligonucleotides to amino acids immobilized or a chromatographic support is expressed by "selectivity coeffi cients" (Saxinger et al.. 1971, Saxinger and Ponnamperuma 197 After Saxinger and Ponnamperuma (1974).

	UpGp	GpUp	ApUp	ApApUp	GpApUp	Ap
gly	10.32	14.05	23.6	63.9	16.6	10
	13.44	14.39	27.5	60.0	19.4	13
trp	95.1	42.1	187.5	2045	60.4	1
	101.3	56.2	177.6	1817	65.4	1

TABLE II-12

Amino acids and nucleotides bound on a synthetic polymer. Adapted from Harpold and Calvin (1973) and Calvin (1975).

Amino acid and nucleotide coupled in reacted compound	% of reacted compound
Phe-A	7.0 (6.7[a])
Gly-A	10.0
Phe-C	2.9
Gly-C	6.5

[a] Calvin (1975).

Using proton magnetic resonance spectroscopy, Raszka an Mandel (1971) studied the interaction of some homopolynucleo with amino acids. Their results were in general agreement w those of Saxinger and Ponnamperuma (section above) and confi (Raszka and Mandel 1972):

> ...the general picture that the 5'-ribonucleoside-P and the corresponding homopolymer interact most strongly with the aromatic amino acids and most weakly with Gl Pro and Lys.

This, in the opinion of Raszka and Mandel, is not encouraging

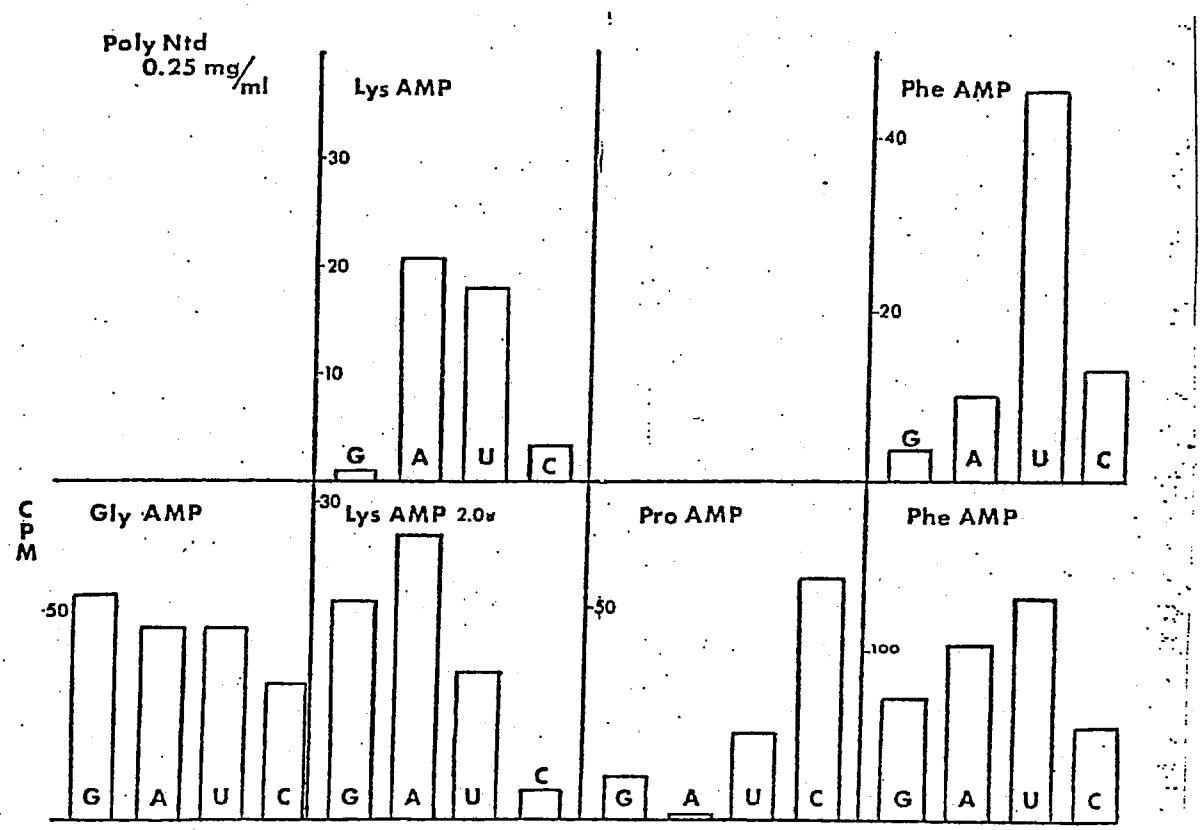
> The experimental results of the binding of amino acids to homopolymer and of the binding of ribonucleoside 5'-monophosphate to immobilized L-amino acids are not consistent with the present codon dictionary in any simple manner. Thus, if one wants to continue to insist on a physical-chemical basis for the genetic code it will require great speculation on the nature of an early primitive code, the proteins it coded for and how it evolved into the present code.

Nakashima and Fox (1972) studied the incorporation of amino acids in particles formed of lysine-rich proteinoid and each of four enzymically synthesized homopolynucleotides. The were able to find empirically a set of conditions in which the amino acids were incorporated preferentially in the microparticles which contained the homopolynucleotides "related" to their present day codons (Fig. II-14).

FIGURE II-14

Incorporation of amino acids from their adenylates into micro particles prepared from lysine rich proteinoids and homopoly-nucleotides under selected conditions.

The figure shows six cases, for four amino acids, as obtained with a polynucleotide concentration of 0.25 mg/ml. The results are "codonic" in the following sense (using Lys-AMP in the top row as an example): about twenty poly(A)-complexed proteinoid microparticles (CPM) incorporated lysine; fewer poly(U)-CPM did so, and very few poly(G)- and poly(C)-CPM; AAA is a codon for lysine. The results of the bottom row were obtained with different amounts of AMP-anhydrides. Changing the concentration of polynucleotide produces irregular results (codonic, anticodonic and of "no evident code-related quality"). After Nakashima and Fox (1972) and Fox et al. (1972).



II.5 Prebiotic Chemistry and the Genetic Code.

The self-assembly of the constituents of living systems or their precursors in the absence of today's complicated genetic system is still a fundamental problem of prebiotic chemistry. Yet, although there is no general agreement on pa ticular prebiotic pathways or initial conditions, nobody doub that if we knew these conditions the problem would be only on of time and/or techniques. One natural extension of this think- ing is that, at least in principle, today's genetic code woul be predictable if we had a deeper understanding of the chemis- try involved.

Do the experiments on the interactions between nucleotide and amino acids justify this assumption, and what alternatives are p

The most common approach to this experimental problem has been to study very simple systems, as exemplified by the follo ing statement (Raszka and Mandel, 1972):

> If the present genetic code is representative of a preferred physical interaction, however weak between amino acid and codon, we should most easily recognize this in a study of the four homopolymers (poly A, poly U, poly G, poly C) and the corres- ponding amino acids they encode.

Their conclusions were pessimistic (previous section) but thei view is not shared by everyone.

Referring to the experiments of Harpold and Calvin (previous section), Calvin (1975) states[1]:

> We thus have the beginning of evidence that, even with one amino acid and one base, there is a kind of selectivity intrinsic in the structures.

Fox (1974) comes to a similar conclusion[1]:

> The molecular preferences are mostly not strong but a stereochemical basis for interaction is evident. Presumably, after the first macromolecular selections initiated the basis for the genetic code, enzymes evolved to magnify the inherent differences.

But the approach of his group differed from that of most experimenters (Nakashima and Fox 1972):

> The conditions necessary to achieve the results reported were arrived at empirically. Constructionistic research of this type has the benefit of being designed for goals that have already been designed by, and shown to have been attainable by, the evolutionary process.

And their results were qualitatively as well as quantitatively sensitive to the experimental conditions (Fox et al. 1972):

> ... a somewhat special range of conditions is required to demonstrate such a correspondence. For instance, in the preparation of the microparticles, the concentration of polynucleotide must be 0.25 mg/ml rather than, for example, 0.75 mg/ml.

---

[1] Cf. also with the quotation by the same author in the introductory chapter.

This could imply that today's code depends on the composition of the primitive soup rather than on a special interaction of codons (anticodons) with amino acids. As I will show in the next chapter, this idea is common to several theories of the origin of the code.

I think we can conclude the following:

1. Conditions on the primitive Earth may have been favourabl to the production of macromolecular systems which could have been the precursors of living systems.
2. Nucleotides and amino acids interact selectively at sever levels: monomer-monomer, polymer-monomer, polymer-polyme
   a) No all-or-none specificity has been observed.
   b) No codon-amino acid assignment could be inferred unam- biguously from the interactions observed.
3. Whether and how nucleotide amino acid interactions helped shap the genetic code remains to be established.

In the light of what certainly is a very complex technical pro lem, the third point is not surprising. There are, however, conceptual aspects of this problem, whose clarification may influence the direction of future experimentation. They are dealt with in the next chapter.

III. OUR PRESENT UNDERSTANDING OF THE GENETIC CODE

III.1 Structure in the Genetic Code

III.1.1 Universality

Evidence from different sources indicates that the genetic code is the same in all contemporary organisms (Ycas 1969, p. 255). Suppressors do cause alterations in codon-amino acid assignments, which may be viewed as exceptions to universality, and some differences in punctuation have also been reported. These and other exceptions, however, are "difficult to interpret" (Woese 1967, p. 164), and it cannot be disputed that "the code has remained constant over a long evolutionary period" (Watson 1970, p. 431).

As soon as evidence for universality became available, it was recognized that the phenomenon was not surprising, in light of the drastic effect - most likely lethal - any change in the code would certainly have, throughout an organism (Crick 1963b, Hinegardner and Engelberg 1963). Unfortunately there are too many good reasons in favor of universality which, as we shall see, can be made a requirement of all types of models proposed for the origin of the code. As Woese pointed out (1967, p. 163) the crucial test of all models will depend instead on their prediction of a nonuniversal element in the code. Clearly work is needed from both directions: refinement of the models to allow for such predictions, and clear demonstrations of nonuniversal aspects.

III.1.2 The Chemical Nature of the Genetic Code

I have already mentioned in section II.4.2 how a "polar requirement" defined purely on the basis of chemical properties of amino acids (Woese *et al*. 1966) reveals some striking regularities in the gentic code: table III-1 shows that codons which differ only in the third position are often assigned to the same amino acid (which had already been observed) or, at least, to amino acids which have similar polar requirements; in addition, codons which have U and C in their second position are assigned respectively to amino acids also with similar polar requirements. Pelc and Walton (1966) observed that

> the amino-acids with hydrocarbon residues have U (uridine) or C(cytosine) as the second base, those with branched methyl groups have U as the second base; the basic and acidic amino-acids have A (adenine) or G (guanine) as the second base, the aromatic amino-acids and amino-acids derived from a common organic acid are grouped together; ... if the third base in a triplet influences the coding, it does so only by being either a purine or a pyrimidine.

Some correlations have also been found, based on the molecular weight of amino acids (Schutzenberger *et al*. 1969).

This type of evidence naturally leads to look for codon-amino acid interactions (section II.4.2), but inasmuch as the chemical properties of amino acids are related to the structure and function of the proteins they form, the association

TABLE III-1

Codons which differ in the third position are assigned to amino acids with similar "polar requirements". After Woese et al. (1966).

UUU		UCU		UAU		UGU	
	Phe 5.0			Tyr 5.4		Cys 4.8
UUC		UCC		UAC		UGC	
			Ser 7.5				
UUA		UCA		UAA		UGA	
	(Leu)						Trp 5.2
UUG		UCG		UAG		UGG	
CUU		CCU		CAU		CGU	
				His 8.4			
CUC		CCC		CAC		CGC	
	Leu 4.9		Pro 6.6				Arg 9.1
CUA		CCA		CAA		CGA	
				Gln 8.6			
CUG		CCG		CAG		CGG	
AUU		ACU		AAU		AGU	
	Ile 4.9			Asn 10.0		(Ser)
AUC		ACC		AAC		AGC	
			Thr 6.6				
AUA	Ile	ACA		AAA		AGA	
				Lys 10.1			(Arg)
AUG	Met 5.3	ACG		AAG		AGG	
GUU		GCU		GAU		GGU	
				Asp 13.0			
GUC		GCC		GAC		GGC	
	Val 5.6		Ala 7.0				Gly 7.9
GUA		GCA		GAA		GGA	
				Glu 12.5			
GUG		GCG		GAG		GGG	

of classes of codons with classes of amino acids can also be interpreted as the result of selection pressure, which would tend to minimize the deleterious effect of mutations in the genome; i.e. misreading of a codon would likely cause the assignment of an amino acid in some way "similar" to the correct one (Sonneborn 1965). Particularly striking in this respect is the observation that the code minimizes the possibility that a random change in a codon base will cause the replacement of a hydrophylic amino acid with a hydrophobic one (Epstein 1966, Volkenstein 1966). Optimization of course implies that the codon catalog evolved. An extensive list of chemical properties which may have contributed to the final form of the code has been compiled recently by Jorre and Curnov (1975) on the basis of statistical calculations.

In conclusion it can certainly be said that the present day codon amino acid assignments are not random, but it is not clear how and at what point chemical constraints helped shape the catalog, which of course is the crux of the whole matter. The major lines of thinking will be discussed in section III.2.

III.1.3 Linguistic Structure of the Genetic Code

> ... the hereditary propagation of a trait involves a description or code and therefore must involve a classification process and not simply the operation of inexorable physical laws of motion on a set of initial conditions. These laws of motion depend only on the immediate past and cannot be directly associated with the concepts of memory, description, code and classification...

(Pattee 1967). A classification process requires that a distinction be made between genotype and phenotype. This, in the opinion of the same author, is a "fundamental evolutionary principle" whose logical aspects were included in Von Neumann's design of a self-replicating automaton, and were pursued by several other authors (for some references see Pattee 1967).

Pattee then carries the same concept a bit further (1972)

> ... hierarchical control in living systems at all levels requires a set of coherent constraints which in some sense creates a symbolic or message content in physical structures, in other words, a set of constraints which establishes a language structure.

And he stresses that this is not just an interesting analogy:

> ... while many biologists more or less metaphorically think of the genetic process as the "language of life" the full necessity of an authentic language system for the very existence of life, which I am proposing, is seldom recognized.

(see also Pattee 1971). A detailed analysis of the consequence of a linguistic structure for the genetic system, with additional evidence from molecular biology, lead him to conclude that (1972) "there is no known physical, chemical, or logical reason why equivalent alternative codes could not occur in principle". Of course many factors may cause the code to be

unique on earth, but its linguistic structure would lead one to expect, for instance, that life on another planet should be based on a different code, just as we would certainly not expect an extraterrestrial civilization to have English as their natural language.

On second thought, one can be a bit more careful in explaining the implications of a linguistic structure. For instance would it be incompatible with finding the same code on another planet? If not, why? And one could still ask: if different, what elements of an extraterrestrial code would be similar to ours (assuming of course the same chemical elements and the same chemical laws)?

I will attempt to deal with some of these questions, and with the problem of the origin and evolution of such a linguistic structure, in the remaining chapters.

At a deeper level one must also consider whether these concepts should lead us to reconsider our understanding of physical processes in general. For thinking in this direction and for further references the interested reader should consult Pattee (1967, 1971, 1972a,b).

III.2 Theories of the Origin of the Genetic Code

Three major ideas have influenced to some extent every conjecture which has been made about the origin of the genetic code. The first two have already been mentione or implied up to this point.

1. The code was caused by steric affinities between codons (or anticodons) and amino acids.
2. The code resulted from a "frozen accident".
3. Particular initial conditions on the primitive Earth, and early biochemistry, had a major effect in shaping the code.

This is certainly an oversimplification of the amount of thinking that has been devoted to the problem of the origin of the code. Some researchers ("Mixed Models", section III.2.4 mention explicitly that more than one of the above views has something to contribute, but the controversy illustrated by t quotations in my introduction show that one important point hasn't been stressed: there is enough experimental evidence and sound reasoning to show that both chemical and stochastic phenomena played an important role; the problem is precisely that of determining how they are related, rather than which aspect is more important. I will begin to treat their relation in chapter IV.

III.2.1 Influence of Inherent Structural Properties

The main advantage of this point of view (also called 'stereochemical' or 'mechanistic' theory) is that it can gene rate models which are testable. Some of the evidence obtaine has been examined in section II.4.2, and, as I pointed out in section II.5, it is considered encouraging by some workers and discouraging by others.

Universality is an obvious result of this theory, which, i a nutshell, states that the present day genetic code is determined by structures of the codons (or the anticodons) and the amino acids. A careful study of these structures, therefore, should give some indication of what assignments can be expected (e.g. Reanny and Ralph 1967, Lacey and Pruitt 1969, Carter and Kraut 1974).

Crick (e.g. 1967b, 1968) has criticized this point of view, mainly for not being founded on sufficient evidence (see also section II.5). But one question I would like to address myself to is whether, or to what extent, a stereochemi- cal approach is compatible with the linguistic structure of the code (section III.1.3). I will consider this point in chapter IV.

III.2.2 Influence of Primitive Biochemistry and Earth Conditions

This type of thinking has given rise to a number of elaborate proposals for the origin and evolution of the code. These proposals are generally based on the authors' general understanding and intuition of biochemical events as they could occur over long periods of time, and on particular models of primordial conditions.

A good example is the work of Jukes (e.g. 1969a,b, 1973a,b, 1974, 1975). Basing his considerations on the structure of the codon catalog and the frequencies of amino acids in present day proteins (e.g. King and Jukes 1969), he suggests that a primitive code may have comprised only 10 of the present day amino acids, although some different ones may have been present and then lost (Jukes 1973a, 1974). In particular arginine is a late "intruder" which may have displaced ornithine (Jukes 1973b) and may have taken away some codons from lysine. The reason, of course, is selective advantage related to the particular functions assumed by the above amino acids in proteins.

However he doesn't think that the code was shaped by an optimization process (section III.1.2):

> ... the only optimization shown by the code is the fact that many of the changes in the third base of codons do not produce changes in amino acids. This feature may be an incidental result of the spatial nature of codon-anticodon pairing rather than an "evolutionary optimization".

And he objects as well to a stereochemical fit:

> A strong objection to the 'stereochemical fit' concept is that the two sets of codons for serine, UCN and AGY, are so dissimilar. Various authors have proposed that there is a relationship between the second base of codons and the chemical properties of the amino acid, and that this relationship has governed the evolution of the code. Such a relationship, in my opinion, is perceptible only for codons with U in the middle position.
>
> -
>
> The fact that five amino acids with hydrophobic side chains all have codons with a middle U may have been shaped by evolution, or it may be a coincidence.

A similar line of thinking has been followed by Mikelsaar (1975), who proposed that the quantitative structure of the genetic code may have been determined by a pre-existing ratio ("archaeorelation") of amino acids; and by Wong (1975):

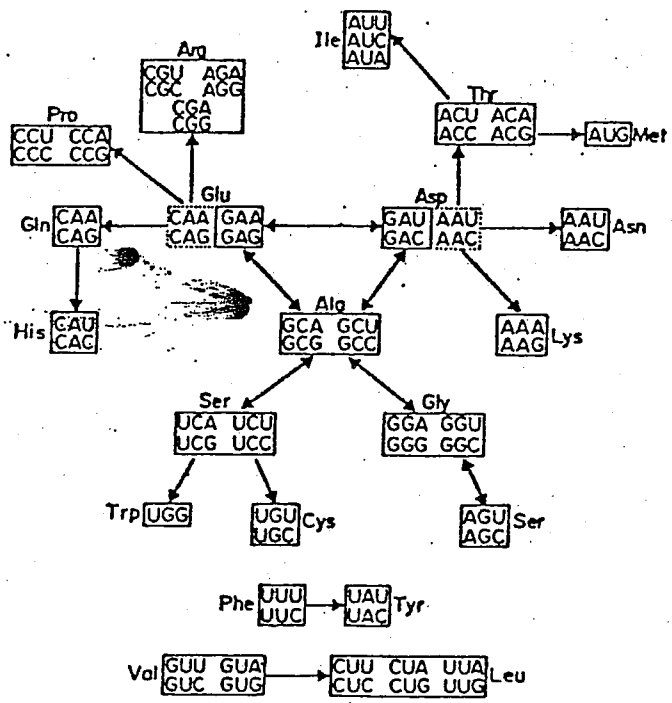
> The structure of the codon system is primarily an imprint of the prebiotic pathways of amino-acid formation, which remain recognizable in the enzymic pathways of amino-acid biosynthesis.

Figure III-1 shows a detailed evolutionary map for the genetic code, as proposed by Wong.

All the above theories imply that there was some element of randomness, especially in the initial choices of codon-amino acid assignments. For other authors instead even the first assignments may have been caused by specific primitive Earth conditions. For instance clays may have played a role (Hartman 1975) or an oil-slick with colloids and micelles (Nagyrary and Fendler 1974).

FIGURE III-1

A "Co-evolution Theory" of the genetic code. Codons used at present are enclosed in boxes. Those in dotted boxes probably belonged to Glu and Asp, as indicated. Amino acids at the tip of the arrows are biosynthetic products of those at the tail (double headed arrows represent interconversions), and occupy contiguous codon domains (single base change). After Wong (1975).



A yet different twist has been proposed by Walker (1974) the present day code is the result of hybridization of two codes which originated separately and whose genomes lacked respectively A and C (or U and G). His considerations were based on a particular type of symmetry in the codon catalog.

III.2.3 The Stochastic Approach

If unique codon-amino acid assignments can only be made in the presence of adaptors, and if adaptors for all possible assignments could in principle be formed in a prebiotic soup, then the formation of a code must wait for the accidental appearance of a correct set of adaptors. The selective advantage a code would give to a system, and the difficulty of changing it without loss of previously stored information cause the code to be "frozen" into its final form; for this reason probabilistic theories of the code are often labeled "frozen accident". Unfortunately the word seems to imply a single event i.e. that a highly complex system was assembled by chance. This certainly goes against everybody's intuition of how complex systems are formed. Interestingly, Crick, who is generally

credited with putting forth the frozen accident idea (e.g. Calvin 1975), envisioned a continuous process of evolution from simpler codes in which chemical affinities may have played an important role, to the final "frozen" form (Crick 1967b, 1968), in which the adaptors play the major role.

One great advantage of dealing with chance is that, after having made some estimates on the type of complexity necessary to keep the system going, one can put some numbers down.

Probability calculations

When attempting to estimate the probability for the origin of life by "pure chance" one is faced immediately with a paradoxical situation: even an average enzyme consisting of 100 amino acids is one of $20^{100} = 10^{130}$ possible structures. The entire universe is estimated to contain about 10^{80} atoms (e.g. Kaplan 1972). Essentially the same situation is at the root of Wigner's paradoxical demonstration that life is incompatible with quantum mechanics (1961).

One can reduce these immense numbers by considering that the search space was limited by chemical resctrictions on amino acid sequences, which caused primitive sequences to be non-random at least to some extent (section II.3.1, Pattee 1961). But the most common approach has been to consider the fact that one really needs only the correct functions rather than specific structures. Of course no one knows how many sequences

can perform the same enzymatic function. Applying these considerations to a set of λ codon-amino acid adaptors Eigen (1971) and Hoffmann (1975) reached slightly different results:

$$P \sim \frac{(\lambda!)^3\,[(\lambda^2-1)!]^2}{[(\lambda^2+\lambda-1)!]^2} \qquad \text{(Eigen)}$$

$$P \sim \frac{(\lambda!)^3}{\lambda^{4\lambda}} \qquad \text{(Hoffmann)}$$

where P is the probability of finding a set of λ adaptors which produce, at the same time, unambiguous assignments and a new set of adaptors which have identical properties. Some values of P are compared in table III-2. Both authors have assumed that a chosen volume element would always contain λ adaptors and that all assignments were equiprobable. Their results combined suggest that a feasible upper limit for λ lies between 6 and 8.

Hoffmann (1975) also generalized his calculations to a system where needed catalytic activities occur without the corresponding disruptive activities. This tends to lower the above probabilities, but, as Hoffmann states:

> The calculation serves simply to illustrate how the problem can be approached, and gives some idea of the relevant orders of magnitude.

And he is still optimistic:

> Although very low probabilities are obtained with feasible values... they are not necessarily too low when compared with the time and space available.

TABLE III-2

The probability (P) of finding a set of λ adaptors which produce, at the same time, unambiguous assignments and a new set of adaptors which have identical properties. The values obtained by Eigen (1971) and Hoffman (1975) are compared. The discrepancy has been acknowledged but not explained by Hoffman.. Adapted from Eigen (1971) and Hoffman (1975).

λ	2	4	8	20
P(Eigen)	2×10^{-2}	1.6×10^{-6}	4×10^{-16}	5×10^{-50}
P(Hoffman)	3×10^{-2}	3×10^{-6}	8×10^{-16}	10^{-49}

FIGURE III-2

Threshold accuracy for the adaptors of a stable translation ap
paratus. The symbols have the following meaning:

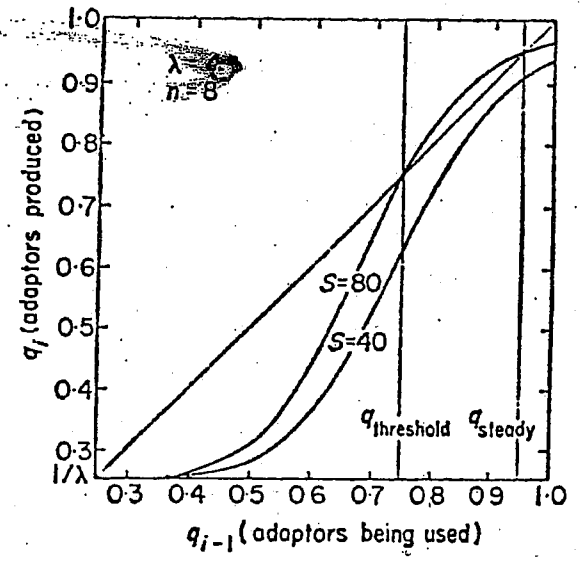
q_i = average accuracy of translation by adaptors of the i^{th} generation (which are therefore produced with q_{i-1} accuracy);

λ = number of different amino acids distinguished

S = specificity of adaptors produced without any errors

n = number of amino acid residues essential for specific substrate recognition (but not essential for adaptor function).

Above $q_{threshold}$ a new generation of adaptors has better average accuracy, until q_{steady} is reached. Below $q_{threshold}$ the accuracy deteriorates. After Hoffmann (1975).



The "error catastrophe" approach

In 1963 Orgel pointed out that since the translation apparatus must be expected to make some errors, these errors would result in defective parts and therefore more errors - the whole system finally collapsing in an "error catastrophe". He suggested that this might be a cause of aging. Later it became clear that the error catastrophe could be avoided (e.g. Orgel 1973), in fact the system, under some conditions, could even rid itself of some errors (Hoffmann 1975, Fig. III-2). This is very attractive from the point of view of the origin of the genetic code, since one expects a primitive system to make more errors than a highly evolved one. With this perspective, the problem of the stability of the translation apparatus is being pursued very actively (e.g. Goel and Yčas 1975, Kirkwood and Halliday 1975).

Recently Goel and Islam (1977) have obtained interesting results by considering the effect of individual synthetases (analogous to Hoffman's adaptors) above or below the threshold level. Both desirable and undesirable effects are possible: (a) just one synthetase above threshold may be sufficient to "correct" the whole remaining set of below theshold synthetases and (b) just one synthetase below threshold can cause all the ones above threshold to deteriorate (error catastrophe).

III.2.4 Mixed Models: Refinement Evolution

Although all but the most extreme models include some of the aspects I have already presented, few authors have made explicit statements to that effect.

I have already had occasion to mention Crick (sections II.4.2, III.2.3). His paper of 1968 deals particularly with how both stereochemical and stochastic effects may have shaped the code; and his hypothesis that few amino acids were coded originally is similar to that of Jukes (section III.2.2).

The translation error (TE) model

This model is "basically but not purely stochastic" (Woese 1967). It envisions a primitive system where only discrimination between classes of amino acids was possible. Thus only "statistical proteins" could be formed. Some of the proteins could act as rudimentary enzymes. To produce better enzymes the translation apparatus had to make fewer errors, but, at the same time, it needed better enzymes to make fewer errors. This looks like the phenomenon which takes place between the threshold point and the steady state in an error catastrophe curve (fig. III-2); the main difference is that here one assumes that whatever the enzymes are doing is the best they can possibly do, given that particular code. In other words the system is in a steady state already. As a way out Woese (1967, p. 158) observed that

> ... a cell would gain (selective advantage) if functional amino acids were assigned to the least error-prone codons, leaving the more error-prone ones to be taken up by the relatively nonfunctional amino acids...

and proposed that, by reshuffling the codon assignments

> The cell would make more and more precise distinctions among the amino acids within the original crude amino acid groups as the process went on and as the enzymatic capabilities of the cell became more sophisticated and precise through gradual improvements in translation.

In the above scheme the effect of point mutations would be minimal at first, to increase gradually as the code becomes more precise. This scheme is designed to overcome a major objection to Sonneborn's lethal mutation model (section III.1.2), i.e. that a search for an optimal code by point mutations would inevitably end up in some local maximum (e.g. Woese 1967, p. 155). The same idea was made more clear by emphasizing (Woese 1970) that the codon assignments were "refined" rather than "reshuffled" by making <u>new distinctions</u> among previously coded groups:

> Codon assignment refinements should <u>not</u> put the cell at <u>any</u> selective disadvantage because the cell did not initially distinguish among the amino acids in the group whose assignment became refined into subgroups. For this reason, a <u>refinement</u> evolution of codon assignments seems theoretically preferable to the conventional <u>reassignment</u> mechanism.

However the fact that refinement evolution would automatically be advantageous to the system is not as clear as implied above.

Yčas (1974) agrees with the idea of lower specificity for a primitive system:

> ... the evidence gives some support to the concept of an earlier biochemical system of lower specificity, having fewer enzymes but catalyzing more reactions and producing more metabolites. It may be noted that while such system is not "simpler" in the sense of having fewer components, it corresponds better to our intuitive notion of a "more primitive" system, one less closely specified and less organized.

But he regards this not as a defect, rather

> ... as something necessary so that a limited number of kinds of cistrons could produce enzymes of sufficiently broad specificity to make a metabolic system possible.

As a consequence

> A mutation producing some loss of ambiguity would, in general, have been lethal, since it would have deleted certain essential enzymic activities.

Probably this question will not be settled until a more quantitative approach will become possible (see also Section V.3).

III.3 Eigen's Theory of Selforganization of Matter

In 1971 Manfred Eigen published a very comprehensive paper on the origin of life. I already mentioned his calculations on the origin of the genetic code in section III.2.3 (stochastic models), but the importance of his work rests, I believe, mainly in his attempt to produce a quantitative formulation of the evolution of matter before the existence of a code.

The assumptions underlying his formulation were greatly clarified in a subsequent article (1973), from which all quotations appearing in the following sections were taken.

III.3.1 The Fundamental Premise

Eigen's fundamental premise is strictly related to his concept of biological information and fitness:

> Information is worthless if it cannot be conserved up to the moment where it is to be read out. Any non-equilibrium structure, subject to thermal motion, will decompose. In particular, the information carrying macromolecular sequences are (and have to be) such non-equilibrium structures. Thus their 'ur-semantics' can only refer to conservation of information, i.e. survival. Advantageous are those sequences which provide further information for improving this quality: information for an enhancement of speed of fidelity of reproduction or protection against decomposition [my emphasis]. Any self-organizing selection mechanism must operate on these properties.

> ... a polymer which manages to achieve a maximum base pairing (preferably involving the more stable G-C pair) therefore will have the best chance of survival....

> The ability of selective self-organization must be associated with peculiar dynamical properties of the material and will be found only under special circumstances.

From the above it seems clear that Eigen identifies informati at the molecular level with a peculiar structural quality which makes a molecular species better able to survive i.e. more fit:

> ... 'fittest' refers to an individual property.... to an optimal combination of reduplication rate, accuracy, and life time [my emphasis].

An important consequence of these statements is that Darwin's principle does not require the existence of "life" (genetic code?) to be applicable:

> The fact that selection is inherent to certain well defined conditions of matter qualifies Darwin's principle as deducible from more fundamental physical laws. The principle then does not just reflect the peculiar historical event of evolution, nor does it require for its application the pre-existence of any form of 'life' (as often has been claimed).

In a nutshell: fitness can be defined for prebioligical stat of matter. For these states three characteristics must be evident:

1. Metabolism ("formation of an intermediate via the consecutive turnover of energy-rich into energy-deficient material").
2. Self-reproduction ("as an inherent property of the intermediates").
3. Mutability ($Q < 1$).

III.3.2 A Phenomenological Representation

> Selection and evolution at the molecular level are processes of fundamentally stochastic nature. They usually start out from a single copy produced in an indeterminate elementary process.... If an advantageous copy appears, it has to reach a certain abundance--depending on the magnitude of its selective advantage--before a deterministic selection is ensured. On the other hand, this indeterminacy then refers essentially to the copy choice, rather than to the process of amplification.

It is the process of amplification, then, that can be describe deterministically. Referring to some chemical entity i "which may consist of a single compound, a macromolecular sequ or even a whole ensemble of interacting species with cooperative reaction behaviour", Eigen defines three phenomenological parameters A_i, Q_i and D_i which refer respectively to reproduction, fidelity of reproduction, and decomposition. He further defines a "selective value"

$$W_i = A_i Q_i - D_i$$

as a "quantity which is decisive in the contest of natural selection". If x_i represents the variation of the population variable, the excess production is given by $\sum_{x=1}^{N} (A_k - D_k) x$ and the average productivity by

$$\overline{E} = \frac{\sum_{k=1}^{N} (A_k - D_k) x_k}{\sum_{k=1}^{N} x_k}$$

Any species ℓ will grow when $W_\ell > \bar{E}$, otherwise it will decay. The decay of species with lower selective value causes a shift of $\bar{E}$ toward the ensemble maximum W_{max}. However, by chance, a new copy may be brought about with $W_i > W_{max}$, thus causing a new round of growth and decay, and a shift of $\bar{E}$ to a new maximum. The result is evolution characterized by a monotonic sequence

$$W_{1m} < W_{2m} < \ldots < W_{opt} .$$

The monotonic behavior generally applies to linear systems and to classes of non-linear reaction systems which

> ... usually include ensembles of species which are linked together via non-linear reaction couplings causing the whole ensemble - after nucleation of the cyclic connection - to behave like a single individual.

This is a highly simplified version of the essential ideas behind Eigen's phenomenological representation of the evolution of matter. It applies quite readily at a full-fledged biological level:

> Asexual multiplication of micro-organisms in a constant medium-requiring regulation of environmental factors and of overall population density - could be described by such a model.

However its application to molecular evolution requires new considerations. With polynucleotides in mind, Eigen relates Q (for some selected species) with "elementary fidelity factors" q_j which can be attributed to each digit which must be

reproduced (complementarity of base pairing is ignored for the moment). Since Q has a minimum value, below which the species must decay, it can be found that there exists a maximum number of digits allowed for safe reproduction:

$$V_{max} = \frac{|\ln Q_{min}|}{1 - \bar{q}}$$

where V represents the number of digits and $\bar{q}$ is the average fidelity factor.

Consideration of base pairing of complementary strands complicates the problem and the notation, but the basic ideas remain unchanged: $W_i = A_i Q_i - D_i$ is replaced by the two eigen values of the matrix

$$\begin{pmatrix} -D_{+i} & A_{+i}\, Q_{+i} \\ A_{-i}\, Q_{-i} & -D_{-1} \end{pmatrix}$$

where + and - refer to complementary strands. Of the two eigenvalues one can be positive, and takes the place of W_i in representing the selective value; the other one is always negative. The signle digit quality factor q_j is replaced by q_{jk}, where j and k refer to digits belonging to complementary strands.

In the next section I will outline the results that Eigen derived from his formulation.

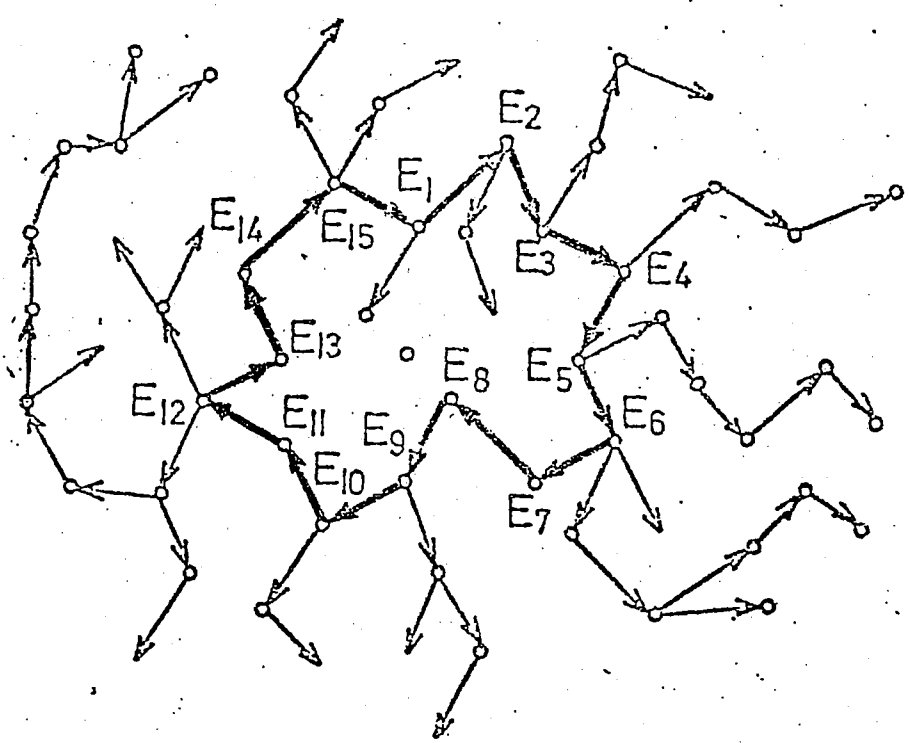
III.3.3 Consequences of the Theory

The most important result, in my opinion, is that the three characteristics listed above are necessary but not sufficient for "an unlimited evolution to the known level of cellular life". This general statement is a consequence of some interesting particular results:

1. "Given realistic values of digit quality factors (q_{ij}), the information capacity that could be accumulated reproducibly in a selected polymeric chain turns out to be much too low to allow for any encoding of an enzyme machinery as would be required for translation and further enhancement of reproduction".
2. An important case in which "the whole ensemble... behave[s] like a single individual" (section III.3.2) is a catalytic network of proteins which includes a closed loop (fig. III-3). Such system has "two shortcomings, which seriously would limit any progree of evolutionary optimization": it grows parasitic branchings, and it woul require a very unlikely coincidence of errors in order to evolve.
3. The next type of cycle considered, called "hypercycle", in an ensemble of mutually reproducing complementary chains which "specifically reinforce each other via a cyclic catalytic coupling". Unfortunately there is "no clear experimental evidence" that nucleic acids could

FIGURE III-3

A catalytic network of proteins which includes a closed loop. Each enzyme in the loop is produced with the help of the preceding enzyme. To some extent this system behaves like a single individual in Eigen's phenomenological representation, but it is a poor candidate for evolutionary optimization. After Eigen (1973).



exert the needed catalytic influences.

4. The last type of cycle considered by Eigen involves both nucleic acids and proteins ("catalytic hypercycle"). However this "requires some intrinsic assignment of nucleotide codons to amino-acids, i.e. the origin of a code and translation" [my emphasis].

According to Eigen's statement it is clear that, beyond his probabil calculations of section III.2.3, a mechanism for the origin of the code is not included in his phenomenological theory. At any rate his premises and results must be compared with those of any prebiological model. I plan to do so in the next chapter.

IV. SELF-CODING: MODEL FOR A LINGUISTIC-MOLECULAR SYSTEM

It should be clear from the preceding chapters, that evolu tion from a primitive molecular milieu to a code occurred in stages which most likely were dominated by different types of phenomena. In particular I propose that there are three phenom which can be studied, at least as a first approximation, as separate problems:

1. non enzymatic template replication of polynucleotides
2. evolution of a polypeptide constructing machinery in the absence of a co
3. evolution of a translating machinery and onset of a code.

Clearly nature did not solve these problems separately and in a neat sequence, but it is possible that one may learn something from such a highly idealized approach.

The selection of these particular phenomena will be partly justified in the course of this chapter, and particular emphasis will be placed on models for the transition from construction in the absence of a code to construction in the presence of a co These models are not intended to simulate the biochemistry of a primitive environment. Rather, they are tools to play games which will hopefully sharpen our intuition on the following points:

1. the transition from a set of arbitrary and possibly contradic tory codon-aminoacid assignments to a self-consistent set (co
2. the effect of assignments which are not, or not completely, arbitrary (chemical bias).
3. compatibility of chemical bias with the linguistic structure of the code.

IV.1 Different Types of Amino Acid-Codon Assignments and Related Terminology.

Concepts such as "code", "function" and "randomness" are encountered in a great variety of contexts. To assure that these and other familiar terms will not have a different meaning for the reader, I define here what I mean by them.

A "code" is a complete and self-consistent set of codon-aminoacid assignments. That is, if 0 and 1 are the codons and A, B the aminoacids, {[0→A] , [1→B]} is a code and {[0→B] , [1→A] is another code. [0→A], [0→B], [1→A] ...etc. will simply be called "assignments" and a generic assignment will be indicated as $[N \rightarrow_z]$. I assume that assignments, at least in principle, can be either mediated by an adaptor system or not. The only kind of adaptor system we know is the mRNA-tRNA-synthetase-ribosome. Assignmen which are not mediated by an adaptor system are not found in organisms, but they are often postulated for prebiotic systems; the stereochemical fit is an assignment with no adaptor. (sectio II.4.2 and III.2.1)

I will refer to the first type of assignments as "functiona and to the second type as "chemical". Both types of assignments could give rise to sequences with some degree of randomness, but different mechanisms would be responsible, such as the presence of competing adaptors for the functional type, or simply lack

of specificity for either the functional or chemical type. Unless otherwise specified "random functional" will refer to the presence of competing adaptors. It should also be noted that the word "adaptor" is generally referred only to the tRNAs and/or the amino-acyl synthetase, therefore a competing adaptor is a tRNA-synthetase system which causes a different amino acid to be associated with the same codon. The reason may be in the structure of the tRNA, or in that of the synthetase, or in both.

IV.2 Separating the Relevant Phenomena

IV.2.1 Replication of Polynucleotide Chains.

At what point did replication of polynucleotide chain become necessary for further evolution? Could it ever be accomplished without specific polymerases?

It seems to me that the first question is related to the behaviour of the whole system under consideration: in this case some protocell and its environment.

The second question is instead strictly related to the chemistry of nucleotide polymerization and base pairing. From the evidence of section II.3.2 it is still hard to conceive of an efficient machinery for the reproduction of long double strands in the absence of enzymes, yet it is very clear that

complementary base pairing is the major physical constraint on the system and that it is not due to the presence of specific protein adaptors. The problem of the origin of the code is thus linked with that of self-reproduction of polynucleotides only in the sense that the latter may indeed be needed with so degree of reliability before convergence to a code is possible. Of course the situation is paradoxical if one requires that the needed degree of reliability can be obtained only after the cod has been reached. But perhaps some reliability (whatever is allowed by base pairing without specific enzymes) can allow for some degree of coding, which in turn could cause the reliabilit to increase, and so on. This is roughly the situation proposed by Woese (section III.2.4): unclear, perhaps even counter-intuitive to some, but at least not obviously paradoxical. That this type of phenomenon can take place has been shown by the results of the error catastrophe approach (section III.2.3) but an important difference is that that approach deals with unstable systems, which will either collapse or improve in a ti span which is much shorter than the one envisioned here. This p will be taken up again in section VI.1.2.

In this chapter I assume that the needed reliability of self-reproduction has been reached (without specific enzymes) and is such that the polynucleotides remain essentially unchang while the system is being simulated. I will discuss again and relax this assumption in chapter V.

IV.2.2 Evolution of a Constructing Machinery in the Absence of a Code

The present day genetic code consists of a set of function assignments , and no competing adaptors are present. Rather than entering the controversy of whether such a situation should be considered a likely or an unlikely event in a primitive environment, I asked the following question: Can we define some specific system from which the gentic system may have evolved, and which has identifiable structure more complex than a "prebiotic soup"

Essentially I observed[1] that for a system which had important analogies with a molecular system, competing functions could not survive together, and that a self-consistent set of functional assignments resulted from inherent instability. On the basis of this observation two distinct types of of models have been proposed within ou group: the fluctuation model and the kinetic model. In section IV.3 I will discuss how they differ, they do however agree that the stage which preceded the code had to see the presence of competing funtions. This implies in my opinion the following important assumption: <u>at some point before the origin of the code there existed a polypeptide constructing machinery which was fairly insensitive to whatever coding functions happened to be present at any time</u>. Otherwise the presence of competing adaptors (leading to some degree of randomness

---

1 Summer project for advancement to candidacy, May 1974, Dept. of Biophysical Sciences, SUNYAB.

in the polypeptides) couldn't be tolerated. Hopefully, as already suggested Crick (1968), primitive transfer, messenger and ribosomal RNAs could form such a system (the problem of their formation is of the same type as that o self-reproduction of polynucleotides), and the coding assignments affected only the synthetases and the protein material of the primitive ribosomes.

It is not too difficult to envision how ribosomal proteins might have tolerated some degree of randomness, since they do n appear to exhibit specific catalytic activities. As our understanding of ribosomal function and structure increases, this question will hopefully be settled. The effect of randomness on the synthetases instead cannot be easily ignored, but, as error catastrophe models have shown (section III.2.3), one can conceive of a threshold level of randomness above which the system reaches a steady state rather than collapsing.

The evolution of a constructing machinery remains a crucial problem, but again, at least to some approximation, one that can be separated from the presence of a code.

The origin of genetic punctuation

A genetic code would be useless without the presence of sta and stop signals. On the other hand start and stop signals on the DNA, even before a genetic code, could produce needed sequences of transfer and ribosomal RNA, while such signals on the mRNA could at least cause the production of proteins of needed dimensions. Again, the numerous protein factors required by a

highly evolved control system probably could not tolerate randc ness, but it is not inconceivable that some punctuation was possibl in the presence of competing adaptors. I have assumed that sta and stop signals could be read by the system before the presenc of a genetic code. This assumption is often made tacitly (for instance when the probability of finding a given number of need adaptors is considered equivalent to the probability of finding a genetic code), since it greatly simplifies all models. It would be interesting to know if punctuation actually had to preced genetic coding.

IV.3 Two Pictures of the System Which Preceded Genetic Coding

Our group has considered two slightly different pictures o molecular system which might have preceded genetic coding. Basically we all assume that polynucleotides could reproduce themselves with some degree of reliability; that random constru tion of proteins could take place, and that competing adaptors were present (and caused the above randomness). One of us, Vahe Bedian, is pursuing a picture in which all possible competi adaptors are present, and in such numbers that a dynamical treatment of the system, via concentrations and rate equations is possible. My picture instead envisions a chemical system in which all possible competing adaptors can be present, with only a few, or even none,

present at any time. Based on this picture, the model that follows enabled me to play simple computer games which refined intuitive understanding of the origin of the genetic code.

IV.4 The Model of a Molecular Pool in Which Random Construction Can Take Place : The Fluctuation Model

"Random construction" here stands for "construction with competing adaptor assignments of amino acids". The model makes use of 2 types of entities: a set G of "descriptors" (g_1, g_2... and a set E of polypeptides (e_1, e_2,...e_n). The transcription step is ignored, therefore G can either represent genes or mRNAs. The codons which make up a particular sequence g_i will be symbolized by numbers, and the amino acids of a particular sequence e_j will be symbolized by capital letters. Each linear sequence of amino acids is assumed to acquire a particular three-dimensional structure and, as a result, some function within the system considered. I am particularly interested in the synthetases which, with the tRNA's, form the part of the adaptor system that "chooses" the amino acids to be assigned to particular codons.

IV.4.1 Self-Coding System: A Simple Abstraction of the Genetic Code

Consider the following system:

descriptors $G = (g_1, g_2)$

codons (0,1)

aminoacids (A,B)

synthetases $E = (e_1, e_2)$

where e_1 and e_2 are defined by particular sequences of amin acids and cause the assignments [0→A] and [1→B] respectivel or [0→B] and [1→A]. We call the system "self-coding" if the underlying constructing mechanism acting on g_1 and g_2 and making use of e_1 and e_2 as synthetases produces only e_1 and e_2. Of course the model can be easily extended to include any number of codons and aminoacids (provided they can form a code), and any number of e's and also g's which code for polypeptide sequences whose resulting structures can, in principle, have any needed biological function.

Example:

g_1 = 000100

g_2 = 001010

and

e_1 = AAABAA

e_2 = AABABA

where e_1 and e_2 are responsible for the assignments [0→A] and [1→B] respectively, and g_1 and g_2 code respectively for e_1 and e_2.

It is of some interest to model also the fact that not al aminoacid substitutions within an enzyme affect its function. I shall simply say that no substitutions are allowed in the underlined part without changing the function previously assig to the sequence. One can therefore define a set of synthetase S(N→z) which includes all synthetases which cause the assign- ment [N→z]. According to the example above S(0→A) and S(1 would include respectively xAABxx and xxxABA, where x = A or B. These synthetases, in the presence of g_1 and g_2, would invariably ca the production of e_1 and e_2. Thus, by my definition, only g_1, g_2 and e_1, e_2 form a self-coding system.

Within this scheme two types of mutations are possible: neutral and lethal. If for instance g_1 mutates to g_3 = 100 g_3 and g_2 code respectively for e_3 = BAABAA and for e_2, still a self-coding system. However, a mutation of g_1 to g_4 010100 causes the production of e_4 = ABABAA which cannot per form the needed function [0→A]. After the decay of e_1 the system would lose its ability to code for itself. It would be very interesting to examine the evolution of systems which in- clude biological functions other than translation, but we are always faced with the problem that functions must be assigned ad hoc to any new sequence that is formed; they would therefore evolve along pathways designed by us.

---

1 By assuming appropriate sequences of amino acids.

IV.4.2 The Competition Game

I would like to understand how a code is obtained from a system which allows all possible assignments. Here is a simple game which illustrates the problem

Let

g_1 = 000100

g_2 = 001010

e_1 = AAABAA [0→A]

e_2 = AABABA [1→B]

and

g_3 = 010100

g_4 = 110011

e_3 = BABABB [0→B]

e_4 = AABBAA [1→A]

be two self-coding systems. We start with G (g_1, g_2, g_3, g_4) and proceed according to these rules:

1. G remains unchanged.
2. The first polypeptides are produced by random generation of sequences.
3. If synthetases are formed they specify the next assignments according to their functions.
4. In the absence of synthetases or when their functions compe random assignments are made.
5. Only the last set of synthetases produced (if any) causes t new assignments.[1]

---

1 It is perhaps opportune to emphasize again that these rules not intended to simulate a primitive environment bur rather provide one with a clear and fast evolving example of how a code can be "locked in" by the internal logics of the system

Examples of tables IV.1a,b

By random assignments the first polypeptides are produced. Let these be (table IV.1a):

BBBBBA BAAAAA AAABBB BBABAA .

On the basis of our definition they have the following functio

none , 0→B , 0→A , none

respectively. We see that the synthetases present only cause random assignments ([0→B] and [0→A] compete).

Again at random then, the next pool of polypeptides is produced

ABABAA AABBBB AABABB ABBBAB .

Now the function [1→A] is present without competitors, so for the next pool A is assigned whenever the g's have a "1" , otherwise a random assignment is made.

The game proceeds along the same lines until either e_1 a e_2 or e_3 and e_4 are produced, in which case the result is "frozen in". The resulting code is {[0→A] , [1→B]} , i.e. IV.1b the resulting code is {[0→B] , [1→A]} . A summary for 1 trials is shown in table IV.2 and the computer program can be found in the appendix.

If the particular set of polypeptides produced with each generation is considered to define a state of a stochasti automaton, one can, in principle calculate the probability of reaching a code after any chosen number of generations. Unfortunately, even for such a simple system, this requires calculating different probabilities for a very large number of pathways through the states of the automaton. For this game the number of pathways is $(2^{24})^n$, where n is the number

TABLES IV.1a, b

Typical runs of the computer game.

Games are represented by sequences of 0's and 1's (codons), and polypeptides are represented by sequences of A's and B's. The genes are fixed and belong to the two possible self-coding systems. The polypeptides are produced according to rules specified in the text, and their functions (if any) are listed after each sequence.

Codes {[0→A] , [1→B]} and {[0→B] , [1→A]} are chosen respectively in tables IV.1a and b .

TABLE IV.1a

Trial NR 98

0 0 0 1 0 0		0 0 1 0 1 0		0 1 0 1 0 0		1 1 0 0 1 1	
B B B B B A		B A A A A A	0→B	A A A B B B	0→A	B B A B A A	
A B A B A A		A A B B B B		A A B A B B		A B B B A B	1→A
B B A A B A	1→B	B A A A A A	0→B	B A A A B A	1→B 0→B	A A A B A A	0→A
A B B B A B	1→A	B A B B B B		B B B B A B	1→A	B B A A B B	
A B B A A B		A B A B A B		A A A A A A		A A A B A A	0→A
A A A A A A		A A A A B A	1→B	A A A A A A		B A A A A B	0→B
B B B B B B		B B B B B B		B B B B B B		B B B B B B	
A B A B B A		A A B A B B		A A A A A B		A B B A B B	
A B B B B B		B A B B A B	1→A	A B A B B B		A B B B B A	
B B A A B B		A B A B A B		B A A A B A	1→B 0→B	A A A B A A	0→A
A A B B B A		A A B B B B		A B A B A A		B B A B B B	
A B B A B A	1→B	A A B A A A		B A B A B A	1 B 0→B	A B B A A A	
B B B B B B		B B B B B B		B B B B B B		B B B B B B	
A A B A B B		B B B B B A		B A A B B B	0→A	B A A A A A	0→B
A A A B A A	0→A	A A A A A B		B B A B B A		B A A B A B	0→A
A A A B A A	0→A	A A B A B A	1→B	A B A B A A		B B A A A B	
A A A B A A	0→A	A A B A B A	1→B	A B A B A A		B B A A B B	

TABLE IV 1b

TRIAL NR 100

0 0 0 1 0 0		0 0 1 0 1 0		0 1 0 1 0 0			1 1 0 0 1 1		
A A B B A B	1→A	B A A B A B	0→A	A A A A B A	1→B		B A A B A A	0→A	
A A A A A A		A A A A A A		A A A B A A	0→A		A A A A A B		
A A A A A A		A A B A B A	1→B	A B A A A A			B A A A B A	1→B	0→B
B B B B B B		B B B B B B		B B B B B B			B B B B B B		
B B B A B B		A B A A A B		B B B B A A	1→A		B B B B A B	1→A	
B B B A B A	1→B	B B A A A B		B A A A A B	0→B		A A B B A A	1→A	
B B B A B B		B B B B A B	1→A	B B B A B B			B A B B B B		
A A B A A B		A B A A A A		B A B A B A	1→B	0→B	A A A B A A	0→A	
B A A B A A	0→A	B B B A B B		A B B B A A	1→A		B B B A B B		
A A A A A A		A A A A A A		A A A A A A			A A A A A A		
A A B B B A		B B A B B B		B A B A B A	1→B	0→B	B B A B A B		
B B B B B B		B B B B B B		B B B B B B			B B B B B B		
A A A B A A	0→A	A B B B A B	1→A	A A A A A B			A A A B A A	0→A	
A A A A A A		A A A A A A		A A A A A A			A A A A A A		
B B A B A A		B A B A B B	0→B	B B B B A A	1→A		A B B B B B		
B B B A B B		B B A B A B		B A B A B B	0→B		A A B B A A	1→A	

TABLE IV.2

Summary for the results of 100 runs. Only the part of the sequence that is crucial to the function is listed.

SUMMARY

FUNCTION	SEQUENCE
0→A	xAABxx
0→B	BAxAxx
1→A	xxBBAx
1→B	xxxABA

CODE 1 = (0→A , 1→B)

CODE 2 = (1→A , 0→B)

In 100 trials CODE 1 was reached 58 times with an average of 9 generations.

CODE 2 was reached 42 times with an average of 8 generations.

of generations.

Notice that, at the end of each game, the parts of the sequences which are not crucial to the functions of the e's are frozen in as well, and that those descriptors which could have coded for the other set of synthetases, codes for a set function less polypeptides. In general we may say that when a self-consistent set of synthetases is formed, the descriptors which are not needed for the set, code for definite polypeptide structures; they have acquired "genetic information" whether or not these structures serve any purpose at first. It is easy to envision how competition between systems of this type, allowing for mutations in the description would eventually favour the appearance of structures which have survival value, within the polypeptides which do not belong to the self-consistent set of synthetases.

IV.5 Observations on the Model

Although the game of section IV.4.2 is rather trivial, it allows a number of observations that should hold for any model which has the same general features described in the introductory remarks of section IV.4 (Fluctuation Model).

IV.5.1 Self-Constraining Ability

Table IV.2 shows that code 1 was reached with an average of 9 generations, code 2 with an average of 8 generations. These results differ from what would have been expected if each generation had been picked at random. The probability[1] of finding one or the other coding set in the absence of constraints would have been $\frac{1}{2^6} + \frac{1}{2^6} \quad \frac{1}{2^{12}} \approx \frac{1}{2^5}$, thus with a mean number of 32 generations.

Another way to see what is happening, is to imagine that the pool of polypeptides is a black box: the synthetase functio are the outputs and the descriptors are the inputs. Let us intr duce descriptors which do not form a self-coding system with any of the needed synthetases: we get a "good" output (say [0→A], [1→B]), on the average, every 32 "bad" ones, but it doesn't stay. If we introduce instead descriptors which form a self-coding system with the needed synthetases, not only will a good output stay on when reached, but it will be reached faster.

Why should this result be general? Because in the presence of its own descriptor a synthetase will increase the chance that an identical synthetase will be produced again. Note that the production of an identical synthetase <u>is guaranteed</u>, by definition, <u>when the whole set is self-coding; it is merely increased when only part of the self-coding set is present</u>. This is due to the logical nature of self-coding rather than to any particular kind of chemical kinetics considered.

---

[1] The probability of finding the crucial parts of the sequences belonging to either one or the other coding set.

IV.5.2 Comparison with Eigen's Assumptions

In our models we may speak of genetic information and its survival value only after a code has been reached. Particular genes (descriptors) would be chosen only for the survival value they impart to the whole system they belong to, and not because they were the victors of a struggle for survival among similar genes. In fact the system must be free to vary both its gene population and its polypeptide population quite freely in order to find self-consistency. If some members of these populations were significantly more stable than others and had been chosen because of this fact, the probability of reaching a code would be greatly decreased. This is because the needed polypeptides and polynucleotides would have to be both optimally stable as "individuals" and, at the same time, members of a self-coding system. The fluctuation model is in contrast with Eigen's concept of fitness at the molecular level (section III.3.1) if this is to be viewed as a major reason for prebiological evolution. A property such as an "optimal combination of reduplication rate, accuracy and life time" would undoubtedly play a role in selecting broad classe of molecules, but, for the reason I gave above, if particular sequences of nucleotides or amino acids were chosen as "fittest individuals", survival of the fittest would prevent evolution rather than explaining it.

IV.5.3 The Linguistic Structure of the Model

Descriptors and polypeptides constitute the geno- and phenotype of the system; the adaptors are responsible for the classification process. The presence of a classification process makes it possible to obtain equivalent alternative codes (section III.1.3).

In the next section I discuss to what extent the above linguistic structure can be retained in the model, and still allow for a preferential choice of some codes.

IV.6 Bias Towards a Particular Code

IV.6.1 Logical Bias

It became clear while playing these games that different codes are not equally likely. This is easy to see in an extreme case: Consider $g = 000$.
It could code for e_1 = AAA or for e_2 = BBB depending on which code wins at the end. Suppose now that the function of AAA is [0→A] while that of BBB is [1→A]. The two functions have an equal initial chance of appearing but, after that, AAA can guarantee its own survival whereas BBB cannot. In the long run this self-perpetuating tendency of [0→A] would make its inclusion in a final code more likely. The nature of this bias is logical rather than chemical and, since it is due to the particular composition of amino acid sequences in crucial adaptor

sites, it appears even when such sequences are all assumed to be of equal length. But although one would not expect large discrepancies between adaptors of similar functions, the number of amino acids which must be strictly specified may be different for different adaptors. Examples of this type of weighting in favour of one code are shown in tables IV.3a,b.

The effect of logical bias

Would the above effect play an important role or a negligibl one in the selection of a particular code? It is clearly a small number effect, but precisely because of this, in a picture of the origin of the code where the first few functional assignments to appear play a major role, logical bias may be important. For instance, I think that these considerations apply to Woese's model (section III.2.4), which will consider in greater detail in Chapter V.

In the picture proposed instead by Bedian (introduction to section IV.3), the system would be much less sensitive to logical bias: since competition is assumed to take place among a mixture of complete sets of adaptors for all codes, any particular order in which adaptors may have been produced within the system, no longer plays a role.

TABLES IV.3a,b

Logical weighting in favour of a code. The amino acid sequence which must be specified is different for different adaptors. Codes which allow more randomness are favoured.

a)

FUNCTION	SEQUENCE
0→A	AAABAx
0→B	BAxAxx
1→A	xxBBAx
1→B	xABABA

Code 1 = (0→A , 1→B)

Code 2 = (1→A , 0→B)

In 100 trials Code 1 was reached 11 times

Code 2 was reached 89 times

b)

FUNCTION	SEQUENCE
0→A	AAABAA
0→B	BAx Axx
1→A	xx BBAx
1→B	AABABA

Code 1 = (0→A , 1→B)

Code 2 = (1→A , 0→B)

In 100 trials Code 1 was reached 4 times

Code 2 was reached 96 times

IV.6.2 Chemical Bias

The fluctuation model has difficulty accomodating a mechanism whereby codon could directly affect (sterechemically) the placement of amino acids in a polypeptide sequence. This is because it assumes the existence of a polypeptide constructing machinery which makes use of adaptors. It is however consistent with stereochemical affinities between anticodons and amino acids. In the system thus envisioned polypeptides would be produced with, at best, a "facilitator" function; later they would evolve into adaptors and assign the appropriate amino acids even without placing them in proximity of the anticodons, as the present day system seems to do (section II.4.2).

However, for a code to be reached in general, the adaptor function must be assumed to be "stronger"; i.e. even if the assignment [1→A]occurred by stereochemical preference, the appearance of an adaptor for [1→B] would cause the latter assignment to be made. The necessity for this will hopefully be made clear in section IV.6.3.

Modeling chemical bias

Chemical bias can be introduced in the computer games simpl by imposing the condition that, in the absence of adaptors, assignments will be made not at random but with weighed probability. For instanc when [1→A] was chosen 75% of the time in the absence of adaptors for [1→A] or [1→B] , code {[1→A] , [0→B]} was reached 71 times in 100 trials.

Increasing the chemical bias towards a certain assignment is found to increase the likelihood that it will be included in the final code, but of course one can never guarantee such outcome 100% if one assumes, as I have, that a competing adaptor can take over if produced by chance.

IV.6.3 Unlikely Structures Vs. Unlikely Laws

The model in this chapter was designed to propose and illustrate the following points:

1. The persistence of a code is due to the logical self-consiste of the system rather than to an energy minimum.
2. Chemical affinities may well have contributed or even have had a major role in determining which assignments became part of today's code, however, at least in principle, they were not needed.

What was needed was the ability to make arbitrary choices (linguistic structure). Consider for example a system of three codons (0,1, and three aminoacids (A,B,C). Without ambiguities only 6 codes are possible (the number of different orderings of three element Chemical affinities instead have no "need" to make self-consist choices. They might well be {[0→A] , [1→A] , [2→C]} or {[0→ [1→B] , [2→A]} or any other. Indeed, they might even coincide with a self-consistent set but that would be an accident. State in another way: competing choices of assignments can be eliminate (i.e. self-consistency reached) if and only if the assignments are arbitrary (i.e. adaptor functions can take over).

Given enough time any structure which is allowed by the laws of physics can be assembled by chance, although "enough time" may well be inconceivably long. But to state that the structure of the genetic code was entirely predetermined by the laws of chemistry, is to shift from the unlikelihood of a particular structure to the unlikelihood of a particular set of la and while nature has time to search for structures under existi laws, it has no possibility of searching for laws.

I wish to re-emphasize, however, that even within an essen tially stochastic model like the one I have presented, some chemical affinities may have played an essential part in deter- mining the choice of a code and possibly, if my intuition is correct, how long it took for the code to come about (the compu games were too fast to show any significant effects). In this sense it is certainly worthwhile to look for a chemical basis t the origin of the code and, as I have emphasised, this is quite compatible with an accidental origin of needed adaptor functions.

The analogy with natural languages is useful at this point: The origin of language must certainly be sought in primitive sounds which conveyed emotional states, or imitated natural noises. Yet we speak of language only when sounds are recognized to have symbolic value, i.e. the assignment of objects to words is essentially arbitrary. While the symbolic value of words is made obvious by the existence of several languages, the symbolic relationship between codons and amino acids has generally

only been inferred from the existence of adaptors and from the assumption that alternative adaptors could have caused assignments which are not found in today's code. On one hand I have weakened this assumption by emphasizing that chemical biases could have played an essential role in the final choice of the code in spite of the presence of adaptors; on the other hand, the final self-consitency of a coding system can only be explained by the presence of adaptors.

Far from being contradictory, "natural" biases and arbitrari
of
are interwoven aspects natural languages as well as of the genetic code; thus any theory that seeks to solve the problem of its origin must deal specifically with the interrelationhip of both these aspects.

V. THE NEED FOR MORE REALISTIC ASSUMPTIONS

V.1 The Assumption of Relative Constancy of the Descriptor Population

Consider again the competition game (section VI.4.2):

1. The set of descriptors was assumed to belong to a self-coding system.
2. The first set of polypeptides was produced at random. Every new set was produced by the previous one by acting on the descriptors.
3. The descriptors were thus assumed to either stay in the system or be reproduced perfectly by some non-enzymic mechanism.

Clearly I chose the best of both worlds. Of course by assuming that the set of descriptors belonged to a self-coding system I already assumed the problem half-solved. By imposing further that they stay around, I could use them as constraints to find the needed set of synthetases faster than I could have otherwise.

The next obvious step is to start from a random set of descriptors as well. One has then the choice of allowing them to vary faster, slower, or at the same rate as the set of polypeptides.

If there are L sets of descriptors and L sets of polypeptides and only one self-coding system, the chance of

finding it would be $\frac{1}{L^2}$ if the descriptors varied at a faster or equal rate, and be somewhere between $\frac{1}{L}$ and $\frac{1}{L^2}$ otherwise. The latter case was illustrated in the examples of chapter IV and restated above; the former can be seen readily if one considers that, even if one is fortunate enough to find a correct descriptor set and to start constructing the correct set of synthetase one loses the correct descriptor set when or before the set of synthetases is completed.

Of course the opposite extreme is just as bad: if the descriptor set never changes no search for a self-coding system is possible either the descriptors belong to the self-coding system from the very beginning or they never will. With these problems in mind,

There is another important reason why the descriptor population should change slowly: For instance a proto-organism which finds itself with a genetic system must maintain it long enough for this system to provide selective advantage. Until the proto-organism can use the system to record and retrieve its successful mutations for its own survival, a genetic system alone has no logical or physical "need" to maintain itself.

To restate the problem in terms of the models of chapter IV:

let 1100 0101

AABB BABA

be a self-coding system where AABB and BABA provide respectively the functions [1 → A] and [0 → B]

Then the next generation of synthetases will also be

AABB BABA

rather than, say, BBAA or any other, because of logical reasons (internal), if the descriptors are kept in the system long enough. But the descriptors can be kept around only by external reasons:

1. enzymes are produced, which can insure the error-free reproduction of the descriptors (but this can only happen after the onset of a code and after some selection),
2. or, as implied at the beginning of this chapter, resistance to decay and/or reproduction with low mutation rate are chemical characteristics of the descriptor population.

Of course one can always postulate that "everything" happened at once, but this is exactly what I am trying to avoid as much as possible. Now I can restate one of my basic assumptions of chapter IV: The population of descriptors is sufficiently stable, because of its own chemical nature, to remain unchanged for a period before and after the onset of a code.

Experiments and theories aimed at making the above assump quantitative would contribute greatly to the solution of the coding problem.

V.2 The Problem of Selecting the "Correct" Genes

Imagine two urns: one full of all possible genes, the other one full of all possible polypeptides. One might think of the problem of finding a genetic code as analogous to a game of chance in which genes and polypeptides are picked from their respective urns in no particular order, and they are checked each time to see if, together, they form a self-coding system (let's call them in this case correct genes and correct enzymes). The stability assumption I restated in the last section implies that finding a code by chance is not analogous to the above picture, rather it is a little easier; if several polypeptides are tested for each gene, once the correct genes have been found the correct enzymes are found more readily.

But the disturbing fact remains that there is no "educated pathway" to the selection of the correct genes. In trying to explain how a complex system came about I said that half of it came about by pure chance and the other half could have been helped along by the first half - hardly an improvement.

Simple systems might originate from fortuitous interactions among their parts (when the existence of these parts is independent of the existence of the system), but complex systems evolve from simpler systems. Some parts and functions may be modified, added or eliminated, and some relationships between old parts and functions and new ones can generally be recognized.

Can these considerations be applied to the genetic system? The error catastrophe approach (section III.2.3) is one example the old functions are the same as the new ones, except a bit mo error prone. Have functions and parts been added or modified more drastically for the genetic code? In the next section I suggest that they have.

V.3 Recognition of Structure and the Evolution of Codon-Amino Acid Assignments

In chapter IV I made the assumption that given a function of the type, say, 1→B there existed a polypeptide structure which, within the adaptor system, was responsible for the assignment of amino acid B to codon 1. I associated such structures with synthetases and defined them by assigning (ad h a particular sequence of amino acids to each one.

1 and 0 were of course symbolic simplification for 64 codo triplets, A and B for 21 amino acids. I recognized the fact tha in the postulated sequences all positions were not equally cruc in the performance of the function, but I did not consider anot very important principle of a structure function relationship. will state it as follows: Recognition of a structure occurs via recognition of a finite number of properties associated with tha structure, and the most primitive recognition can distinguish t

least detailed properties. We only need to recognize a trunk an branches in order to recognize a tree, if we add the shape of th leaves we can also tell a maple from an oak.

V.3.1 Amino Acid Recognition

Translated at the molecular level, the above principle implies that amino acids are recognized via specific properties (for instan polarity, type of side chains, presence of specific atoms, etc.), and that different parts of the synthetase will recognize (measure) different quantities although some overlap will certainly be possible for some cases. But different properties will require the specification of different numbers of crucial positions (whether or not overlapping), so, within a system of randomly produced polypeptides, structures which could recognize all needed properties together would be much more rare than, say, those which could recognize only one. In other words, it is most likely that distinction between classes of amino acids had to precede distinction between the amino acids themselves, provided that distinction among classes was sufficient to form stable systems.

V.3.2 Codon Recognition

Recognition of the codon to be associated to a particular class of amino acids presents a special problem. Abstractly one can think of the 64 codons as separate objects and apply to them the same considerations made above. There are good reasons to believe that the reading frame may always have consisted at 3 nucleotides (Crick,1968 ; Eigen, 1971), in which case it can be envisioned how a "reading molecule" might find it easier to recognize only the middle nucleotide than the first two together, or all three. The trouble is that the codon is really recognized by base-pairing with the anticodon. At this point however one can either apply the codon-recognition considerations to the anticodon or something similar to the amino acid-recognition considerations, to the entire tRNA structure. The former approach lacks the support of experimental evidence (section II.4.2) but can perhaps be postulated for primitive systems. The latter seems to agree with the way the synthetase is known to interact with the tRNA, but it cannot explain the different role played by the second, first and third nucleotide in the triplet, in determining the properties of the amino acids.

In spite of these unsolved questions, the possibility of relating parts of the triplet with classes of amino acids has some very interesting consequences, as I will try to illustrate below.

V.3.3 Partly Self-Coding Systems

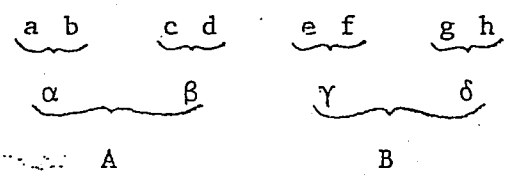
Consider the following system:

nucleotides: (1,0)

codons : (000, 001, 010, 011, 100, 101, 110, 111)

amino acids: (a,b,c,d,e,f,g,h).

Ignoring the problems expressed in section V.3.2, I assume now that the adaptor system never changed the size of the reading frame, but could only read x0x and x1x at first, then 00x , 10x , 01x and 11x, and finally the whole triplets. The amino acids could be lumped into classes as follows:



In analogy with the model of chapter IV I assume further that for every possible assignment there is a set of synthetases that is responsible for that function. I will indicate such a set with S(N→x). For instance S(x0x→A) stands for the set of synthetases which cause the assignment [x0x→A] .

Formally, so far, nothing appears to be very different from what I did in section IV.4.1. But a closer look is needed at this point. Surely it is possible to imagine that, given a certain function, a specific structure will either perform it or not, and that generally there will be more than one structure which can perform a certain function. To say, as I did in section IV.4.1, that only a specific part of that structure is what is necessary for the function, assumes a highly idealized situation

but enables one to specify the whole set of structures which can perform the function. The important point is that at least part of the structure can be specified, and thus be judged fit or not fit to do what is needed.

The situation is different now. The synthetases belonging to the set S(x0x→A), in the presence of any descriptor, cannot specify any structure, and not even part of a structure. They could, say, specify "AB", which represents classes αγ, αδ, βγ, and βδ, thus any of the following amino acid sequences: ae, af, bf, ag, ah, bg, bh, ce, cf, de, df, cg, ch, dg, dh.

I have assumed so far that, due to folding, there exists a one to one mapping between sequences of amino acids and polype structures. In turn, these structures may or may not perform nee function. For instance , a sequence aegbe may or may not form a structure with a needed function, but ABAA specifies several structure, because it represents a class of sequences; and the functions these structures perform are not necessarily related with each other.

I call the set of all polypeptides restricted by a conditio ABA... a "word" and I will represent it as $w_i(N \to x, M \to y)$, where i indicates the descriptor g_i. For instance the descriptor

g_1 = 100 011 001 100

in the presence of S(x0x→A) and S(x1x→B), produces the word

$w_1(x0^{x\to A}\ ,\ x1^{x\to B}) = ABAA$

which represents 4^4 polypeptides. If we are looking for a specific function, say $[x0^{x\to A}]$, the best we can hope for is that at least some of those polypeptides will be able to perform it. I define the set of these polypeptides as $e_i(N\to z)$.

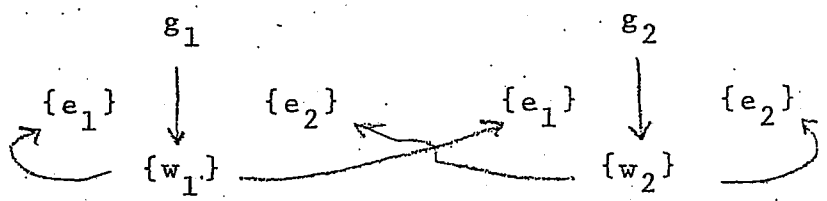
Then

$e_1(x0^{x\to A}) = S(x0^{x\to A}) \cap w_1(x0^{x\to A},\ x1^{x\to B})$,

that is, $e_1(x0x\to A)$ represents the set of polypeptides which perform the function $[x0x\to A]$, and are produced by functions $[x0x\to A]$ and $[x1x\to B]$ acting on the template g_1. Many synthetases could perform the same function, but only those belonging to $e_1(x0x\to A)$ are self-coding in the presence of g_1. Analogously, if an appropr: g_2 exists,

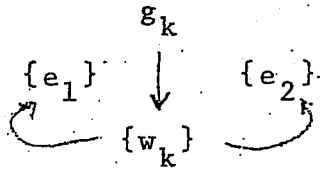
$e_2(x1^{x\to B}) = S(x1^{x\to B}) \cap w_2(x0x\to A,\ x1x\to B)$.

A typical self-coding system which could be formed by these elements would be of the type



See also figure V.1a .

One could even envision a system (fig. V.1b)



where both $\{e_1\}$ and $\{e_2\}$ are subsets of the same word, i.e.

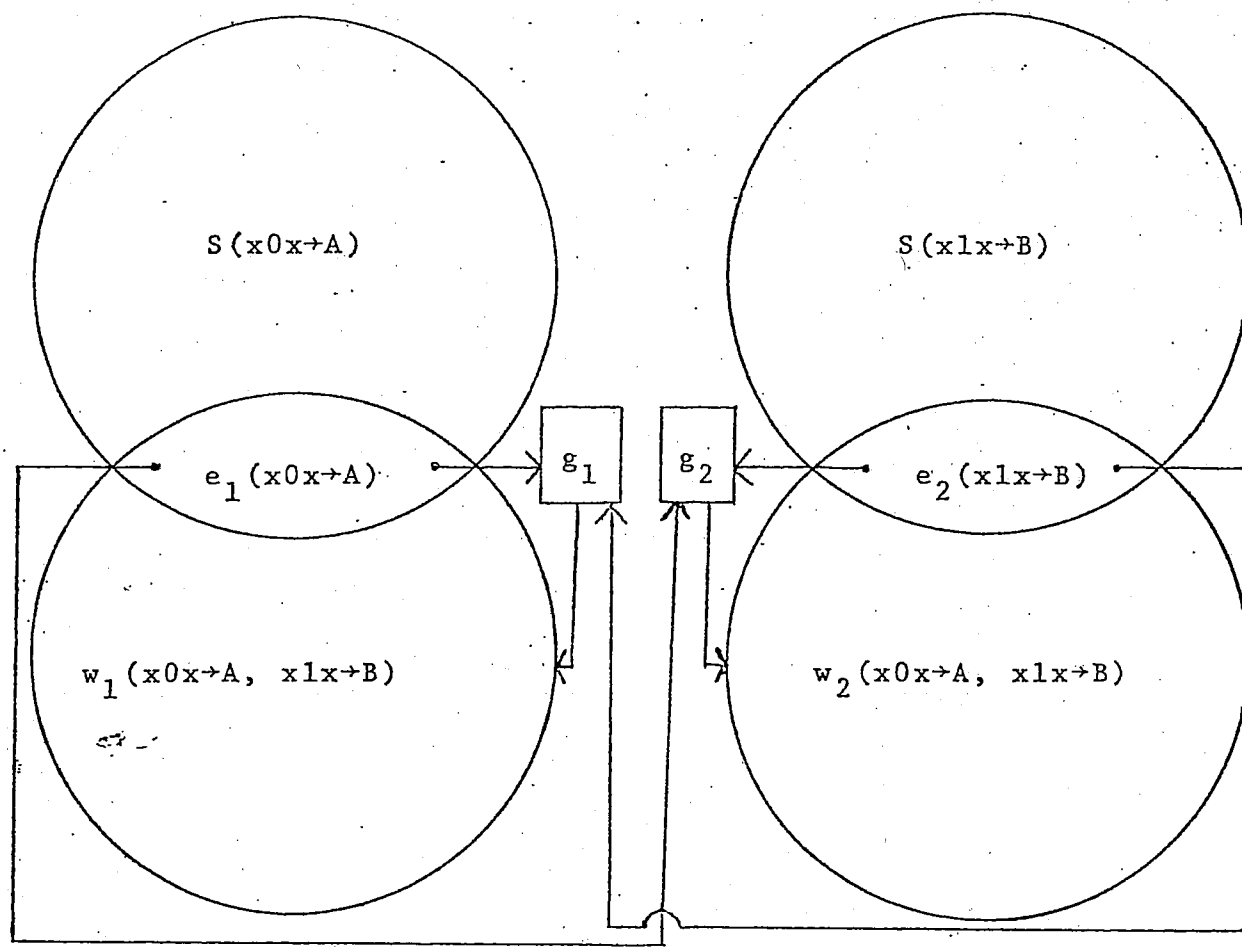
$e_1(x0x\to A) = S(x0x\to A) \cap w_k(x0x\to A,\ x1x\to B)$

$e_2(x1x\to B) = S(x1x\to B) \cap w_k(x0x\to A,\ x1x\to B)$.

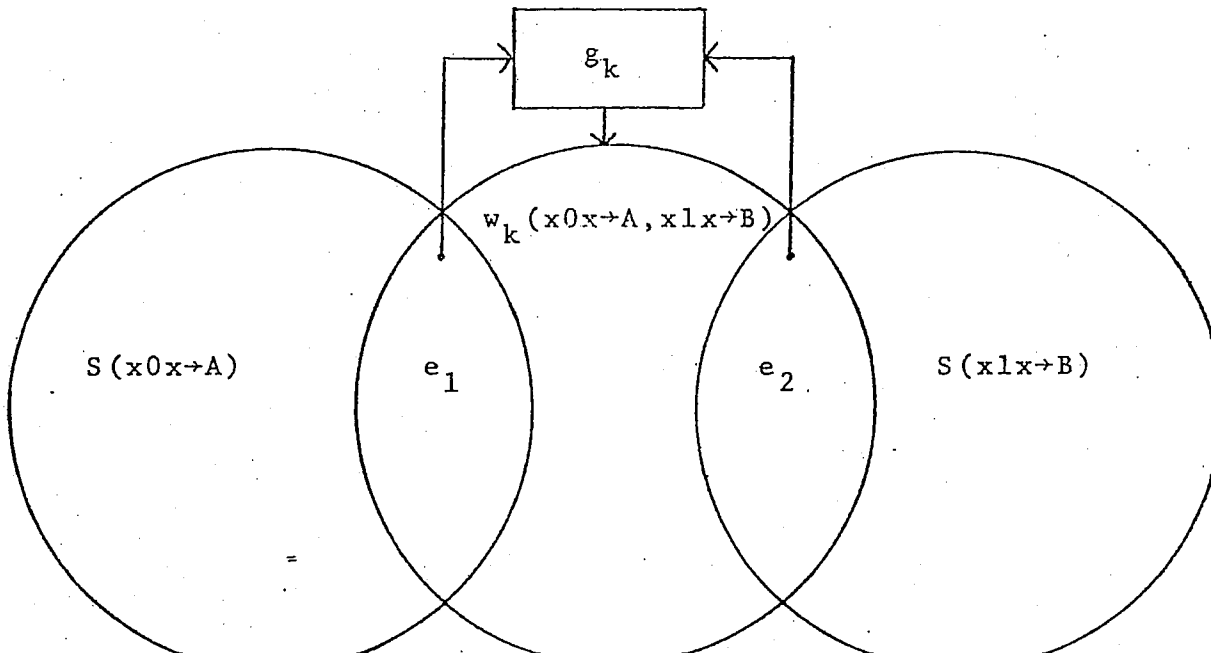
FIGURE V.1a,b

Partly self-coding systems

a)



b)



The important point is that such systems in general could only be <u>partly self-coding</u>, i.e. any needed $\{e_i\}$ could only be a part of one or more $\{w_k\}$ sets produced. The production of some useless polypeptides cannot be avoided.

An explicit example might be helpful:

In the presence of functions [x0x→A] and [x1x→B], gene g_1 = 101 001 010 produces a set AAB of polypeptides. This can divided into subsets in terms of classes α, β, γ, δ:

αaγ

αaδ

αβγ

αβδ

βaγ

βaδ

ββγ

ββδ

which finally represents 64 polypeptide structures, starting with

aae

aaf

aag

aah

abe

abf

... } (58 polypeptides)

...

...

I have called this set w_1 (x0x→A, x1x→B).

We may now list all the polypeptides which cause the assignme
[x0x A], i.e. the set S(x0x→A). Suppose they are

eaf

gbf

aag

hah

abe

then only "aag" and "abe" belong to both w_1(x0x→A,x1x→B) and S(x0x A), thus the set e_1(x0x→A) consists only of "aag" and "abe".

If there is no member of S(x1x→B) among the members of w_1(x0x→A, x1x→B) another gene is needed (example of fig, V-1a). If instead some members of w_1(x0x→A, x1x→B) belong also to S(x1x→B), say "abf", then a second gene is not needed.

V.3.4 Different Levels of Partial Self-Coding. Transitions Between Levels

In analogy with the definition of S(x0x→A) and S(x1x→B), one can define for instance S(00x→α) and S(10x→β), etc.

Clearly

$S(00x\rightarrow\alpha) \subset S(x0x\rightarrow A)$,

$S(10x\rightarrow\beta) \subset S(x0x\rightarrow A)$

and

$$w_1(00x\to\alpha,\quad 10x\to\beta,\quad x1x\to B) \subset w_1(x0x\to A,\ x1x\to B)$$

As before, one needs sets of self-coding structures such as

$$\{e_{11}\} = S(00x\to\alpha) \cap w_1(00x\to\alpha\ ,\ 10x\to\beta\ ,\ x1x\to B)$$

and

$$\{e_{12}\} = S(10x\to\beta) \cap w_1(00x\to\alpha\ ,\ 10x\to\beta\ ,\ x1x\to B)$$

where both $\{e_{11}\}$ and $\{e_{12}\}$ are subsets of $e_1(x0x\to A)$ and the meaning of the second indexes is defined by the expressions on the right. Again there is the possibility that such sets could be empty. If this is the case one must wait for a g_k for which $\{e_{k1}\}$ and $\{e_{k2}\}$ are not empty, but even then, evolution into a system of higher discriminating ability is not an obvious result as I shall try to illustrate below.

Consider, for simplicity, one set $\{e_{k\ell}\}$, and define the ratio

$$\rho(k\ell/k) = \frac{\text{number of structures belonging to } \{e_{k\ell}\}}{\text{number of structures belonging to } \{e_k\}}.$$

Given some gene g_k the distribution of polypeptides belonging to $w_k(x0x\to A,\ x1x\to B)$ will be different, i.e. include more types of polypeptides, than that of $w_k(00x\to\alpha,\ 10x\to\beta,\ x1x\to B)$. For instance assignment $[10x\to\alpha]$ would be excluded from the latter but not

from the former. If the set $\{e_{k\ell}\}$ is not empty, when its members appear - i.e. when $\rho(k\ell/k) \neq 0$ - the distribution of polypeptides in w_k(x0x→A, x1x→B) is shifted towards that of w_k(00→α, 10→β, x1x →B), therefore fewer types of polypeptides are produced. Among these there will be either fewer or more members of $\{e_{k\ell}\}$, and as a consequence, $\rho(k\ell/k)$ will either decrease or increase (fig. V-2a). For reaching a level of higher discriminating ability this feedback loop must of course be positive. This must also be the case when all components are taken into consideration (fig. V-2b).

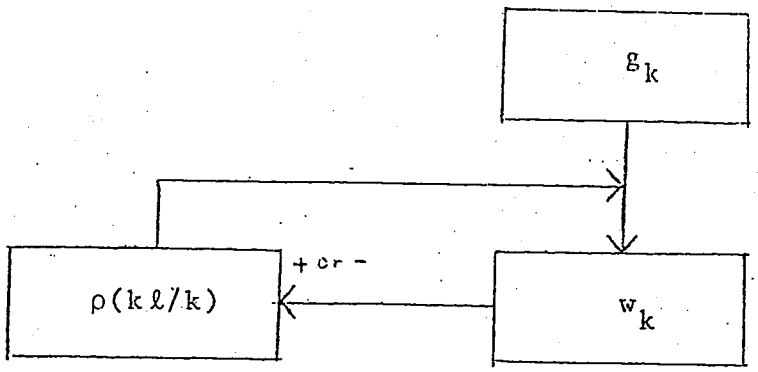
In the total picture one must allow for positive feedback for the assignments belonging to a self-consistent set, such as for instance [00.→α], [10.→β], [01.→γ] and [11.→δ], and negative feedback for the possible competing assignments.

The problem is to determine that there <u>exists</u> a state of the system with the needed positive feedback loops. This cannot be done without introducing more specific assumption and starting a new major effort.

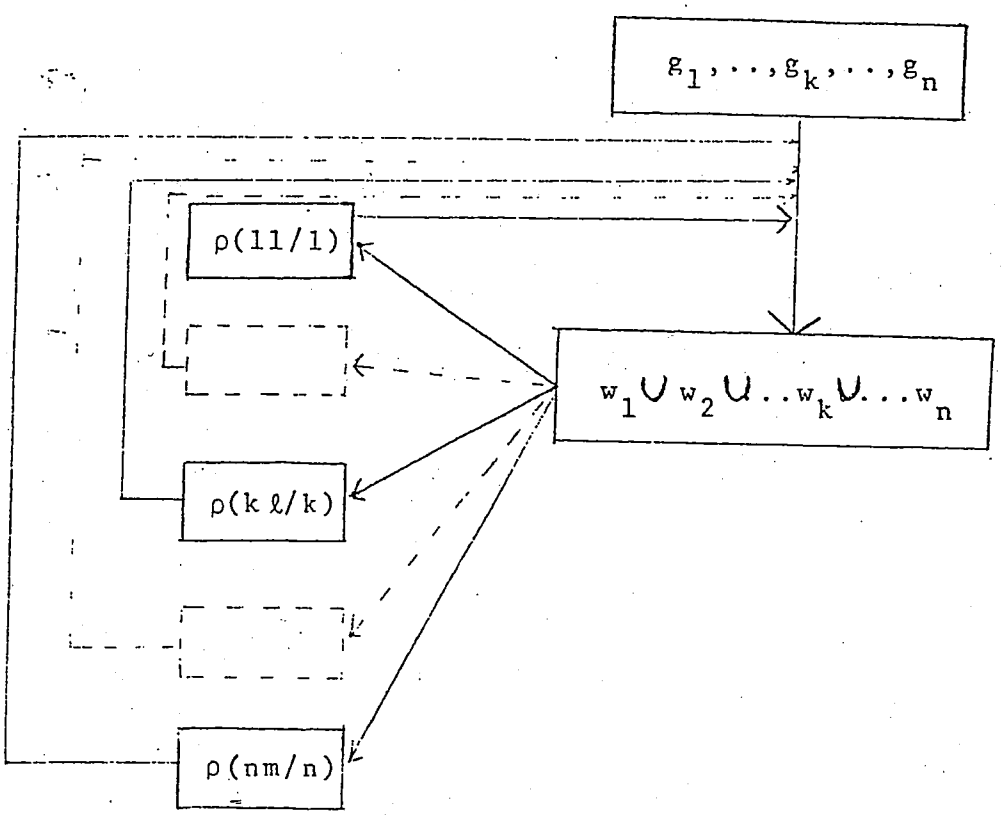
FIGURE V-2a,b

Influence diagrams for $\rho(k\ell/k)$, g_k and w_k. As $\rho(k\ell/k)$ increses the polypeptide distribution in w_k approaches that of the higher discriminating level, which in turn will either increase or decre $\rho(k\ell/k)$.

a)



b)



V.4 A Pathway to the Genetic Code

In section V.2 I mentioned the disturbing fact that there is no "educated pathway" to the selection of the necessary genes. If intermediate states are possible, as envisoned above they are characterized by smaller and smaller sets of polypeptides produced per gene, due to decreased randomness in codon-amino acid assignments. The corresponding decrease in the load of useless polypeptides would certainly help in the selection of such systems.[1]

The decreased randomness of each stage of discriminating ability can be thought of as the elimination of certain regions from the search space. This is due to the fact that once a choice like, say, $[_x0_x \rightarrow A]$ and $[_x 1_x \rightarrow B]$ is made at some stage, the codo 10. can later be assigned either to α or β, but not to γ or δ.

The importance of this evolution scheme is twofold: a) initial primitive molecular systems do not need to be very selective in their search for the genes which belong to a partly self-coding system (mutations where the x's are indicated, are ignored until the system becomes sensitive to that part of the reading frame); later the search would be conducted in a space which has already been restricted; b) competition occurs only among the next possible set of assignments and not among all possible assignments; which

---

1 The objection raised by Yčas (section III.2.4) that such decrease might not always be advantageous, willbe dealt with in the next chapter.

which further invalidated the fluctuation model of chapter IV, whe I assumed instead that competition occurred among all possible final assignments. The last effect could perhaps account for the chemical structure observed in the genetic code (section III.1.2).

In conclusion, this scheme implies that one sould consider th possibility of evolutionary stages in which "living" systems had a "fuzzy" genetic memory of their own structure; i.e. genes could not be used to produce specific polypeptides, but only statistical distributions of sequences with narrower distribution of function. Such scheme follows naturally from the recognition principle applied to molecular structure-function (section V.3), and recaptu some evolutionary continuity within a stochastic approach to the problem of the origin of the genetic code.

VI. CONCLUSIONS AND DIRECTIONS FOR FURTHER WORK

VI.1 An Integrative Approach

The major assumption of this work was that the transition from a pool of polynucleotides and polypeptides with random sequences, could be treated separately from the problems of a) the replication of polynucleotides, b) the evolution of a polypeptide constructing machinery. Based on this assumption I introduced the fluctuation model, and, with illustrative computer games, I tried to clarify the merits and limitations of both stereochemical and stochastic approaches to the problem of the origin of the genetic code. I concluded that some aspects of both approaches are complementary and essential to a theory of the origin of the code. Below I review the major points.

VI.1.1 The Role of Stereochemistry

The fluctuation model assumed that a system of arbitrary adaptors was necessary and would in principle be sufficient for convergence to a code. But it is also clear that the code reflect (although does not require) some undefined "external" chemical reason wh, some codon-amino acid assignments may have been favoured. "External" implies that it does not depend on the adaptor system. One example I can conceive of, for such external

and anticodons. However there is no evidence that this type of mechanism plays any direct role in today's particular code, and mechanisms which include direct codon-amino acid relationsh seem even more unlikely to have had any influence on the final form of the code. There is also no doubt that, in a structural sense, adaptors, codons and amino acids require chemical laws and constraints.

There are two main points I want to make. I think it is necessary to try to understand the role of stereochemical affin as well as of the chemical initial conditions, in determining the early stages of self-organization of prebiotic pools. It may well be that without these stages of self-organization no genetic code would have ever been reached. The analogy with natural languages (Introduction) should again be illuminating. I also think that it would be misleading to try to "find" the code in these affinities, for, as I pointed out in section IV.6. this only raises another problem of why the laws of chemistry should provide us with a self-consistent set of assignments. Self-consistency can be understood much more readily as a result of the presence of systems of adaptors, and it demonstrates that the system is now able to overcome some old constraints (e.g. poorly selective stereochemical affinities) by virtue of some newly acquired ones (e.g. adaptors).

A change to new, more "powerful" constraints is certainly not unique to a genetic system. After all the human primate can now travel at 55 miles per hour and fly well above the eagl What is peculiar about genetic constraints is that they allow

the system to acquire structural information about itself, in fact, as I illustrated in section IV.4.2, it is these constrain that define it. The concept of "structural genetic informati should be kept clearly distinct from the information an observe may choose to gain about a system. The latter depends on the methods of observation and the needs of the observer, the forme depends on the existence of appropriate constraints within the system itself and is independent of the observer. Perhaps it should be called "self-information".

VI.1.2. The Limits of the Stochastic Approach

My approach to the problem of the origin of the genetic code has been fundamentally stochastic. The model I presented in 1974, elaborated further in section IV.4 , is very similar to those used to model the error catastrophe (section III.2.3). This last approach, which perhaps should more appropriately be called "self-correction catastrophe" is fully consistent with this work, and reflects the need to allow for some continuity in chemical evolution. I am particularly aware of that need; but for the very same reason I view the phenomenon of the existe of a threshold for self-correction as a second order requirement in the evolution of the needed structures and functions. Observa of systems which have evolved, either in biology of technology, show that an ancestry can generally be traced for systems, and that each ancestor had the structures that it needed for its own stability and survival. In other words, at each stage of evolutic

all structure-function relations are, in some sense, stable and optimal for the system under consideration. Self-correction shows the transition into a specific stable state from its unst neighborhood. I think we need to better understand the transiti from one stable set of structure-functions to another.

Are there general rules for predicting the evolution of systems, i.e. of their internal structures and functions? That fitness is not a sufficient criterion should be obvious when one considers that all biological systems in existence, from th simplest to the most complex, are, by definition, equally fit. Yet we can establish an evolutionary hierarchy among them. The question can be stated in a different way: one may view organis as producing, occasionally, new versions of themselves, version that are either accepted (fit) or rejected by the environment. We need to say something more about the set of possible version A particular version may have been chosen simply because it came about first and not because it was inherently more viable. Perha this was the case with the code.

I am convinced that a statistical approach which ignores a priori the existence of stable systems which preceded the gene code is not likely to give a satisfactory explanation of its existence. But what do I mean by "satisfactory"? Why would it no be a satisfactory explanation to find that the likelihood of the genetic system simply "coming together" is acceptable, consideri the age of the earth? One reason is that such a theory would not be able to predict anything about the phenomenon of life

except for environments which are very similar to the earth. Another, more speculative, drawback of purely stochastical mode is that with them we might miss the opportunity to understand some important principles which guide the self-organization of matter.

A good example of self-organizing system is provided, agai by language: from sounds to words and sintax, to ideas which require strings of words held together by the rules of sintax, to math and computer languages. At each level the influence of the preceding level is evident, yet entirely new structures app It is evident that to understand these structures one must unde the constraints that caused them.

Consideration of the continuity problem for the genetic code is a natural outcome of this view.

VI.2 The Continuity Problem for the Genetic Code

On the basis of the recognition principle applied to molecular structure funtion (section V.3) I proposed that successive sets of partly self-coding systems preceded the genet code. A theory of the origin of the code should first determine if, and how many such states are indeed possible; after that, determining the probabilities of transition will probably be the the only possible strategy.

Woese's T.E. model (section III.2.4) is qualitatively ident to the evolution model I proposed. The fact that I reached simil

conclusions from a different starting point (the recognition principle as opposed to the necessity of minimizing translation errors) adds strength to the proposal. In addition, I introduced a formalism to describe the behaviour of the system, and pointed to the difficulty of having to predict transitions between stable states.

From the point of view of the self-information of the system it is interesting to notice that its value, calculated for some convenient unit length of the genome, would increase in quantum jumps as the partial codes approach the final genetic code. In other word, the system acquires information about itself in small steps. It might be interesting to speculate on the sgnificance of the similarities between that process and human learning.

Ycas (section III.2.4) criticized the assumption that a transition to a code of better discriminating ability would automatically be advantageous. He pointed out that the system might rely on poor discrimination for the production of some needed enzymes. If a transition occurred at this point the long term beneficial effect might be offset by short term disadvantage. One might argue that the transition would occur for systems where reliance on poor discrimination has been minimized or eliminated, but it is probably fruitless to argue in such general terms. The facts are that evolution in both natural and artificial (technical) systems has increased specialization of function and has brought about finer structural resolution and higher specificity.

At any rate, for quantitative predictions one needs to work with more specific models, and until reasonable models can be constructed these arguments will never be fully satisfactory.

VI.3 Directions for Further Work

I have not developed a quantitative model of the origin of the genetic code, and I can claim no priority for most of the arguments I used. However, I think my "integrative approach" has clarified and exhausted the reasons for the conflict illustrated in the introduction. Thus a new needed perspective has been added, with several specific problems.

In section IV.2.2 I expressed the idea that some genetic control may have been needed for the onset of a code. This implies that the establishment of a control network is another self-organization phenomenon in which a new type of symbolic (?) behaviour is originated: initiation and ending of reading, turning the genes "on" and "off". In other words the first control network has no "purpose" until the genes can be expressed into specific enzymes, or statistical distributions of enzymes. This idea should be checked against all available information on genetic control. It might then be possible to simulate the self-organization of some form of genetic network in the absence of a genetic code.

To be made more quantitative the evolution model should include all available data on recognition of amino acids by synthetases. The crucial point is to determine the difference in structure between an enzyme which can recognize a specific amino acid, and one which can only recognize one or more qualities which characterize it, in other words, some class of amino acids. In section V.3 I assumed that the less specific an enzyme has to be, the more randomness is allowed in its structure (consequently its appearance as a random product is more likely). Studies in folding theory and structure-function relationships are needed to verify this assumption and to find a quantitative relationship between the degree of randomness allowed in the enzyme and its ability to make discriminations.

The evolution model also left out rates, energy, details on the polypeptide constructing machinery and on the replication of polynucleotides. Finally, it would have to include the effect of cellular boundaries and cellular reproduction, which were certainly essential to the evolution of the systems proposed.

Concomitantly with the addition of all the needed details it would be useful to develop a mathematical formalization of partial self-coding, probably along the lines of stochastic automata theory.

In conclusion I hope that this work has at least succeeded in bringing into focus some important aspects of the problem of the origin of the genetic code. It has certainly succeeded in making me realize that I have barely begun.

REFERENCES

Alff-Steinberger, C. (1969) The genetic code and error transmission. Proc.Natl.Acad.Sci. USA 64,584

Allen, G. (1957) Reflexive catalysis, a possible mechanism of molecular duplication in prebiological evolution. Am. Naturalist 91,65.

Andini, S., Benedetti, E., Ferrara, L., Paolillo, L. and Temussi, A.P. (1975) NMR studies of prebiotic polypeptides. Origin Life 6,147

Anker, H.S. (1961) On the geogenous evolution of self-reproducing systems and macromolecules. Perspectives Biol. Med. 5,86.

Banda, P.W. and Ponnamperuma, C. (1971) Polypeptides from the condensation of amino acid adenylates. Space Life Sci. 3,54.

Beck, A., Lohrman, R. and Orgel, L.E. (1967) Phosphorylation with inorganic phosphates at moderate temperatures. Science 157,952.

Bishop, M.J., Lohrmann, R. and Orgel, L.E. (1972) Prebiotic phosphorylation of thymidine at 65° in simulated desert conditions. Nature 237,162.

Burton, F.G. and Neuman, W.F. (1971) On the possible role of crystals in the origins of life: V. The polymerization of glycine. Curr. Mod. Biol. 4,47.

Cairns-Smith, A.G. (1975) Ambiguity in the interpretation of abiotic syntheses. Origins Life 6,265.

Calvin, M. (1975) Chemical evolution. American Scientist 63,169

Carter, C.W. Jr. and Kraut, J. (1974) A proposed model for interaction of polypeptides with RNA. Proc.Natl.Acad. Sci. USA 71,283.

Chada, M.S., Replogle, L., Flores, J. and Ponnamperuma, C. (1971) Possible role of amino acetonitrile in chemical evolution. Bioorg. Chem. 1,369.

Chang, S., Flores, J. and Ponnamperuma, C. (1969) Peptide formation mediated by hydrogen cyanide tetramer: A possible prebiotic process. Proc.Natl.Acad.Sci. USA 64, 1011.

Chang, S., Williams, J.A., Ponnamperuma, C. and Rabinowitz, J. (1970) Phosphorylation of uridine with inorganic phosphates. Space Life Sci. 2,144.

Chapeville, F. (1962) Oxidation of cysteine to cysteic acid on sRNA with retention of specificity in amino acid incorporation. Federation Proc. 21,616d.

Chapeville, F., Lipmann, F., von Ehrenstein, G., Weisblum, B., Roy, W.J. and Bensen, S. (1962) On the role of soluble ribonucleic acid in coding for amino acids. Proc.Natl. Acad.Sci. USA 48,1086.

Chapeville, F., Cartouzou, G. and Lissitzky, S. (1963) Sur la biosynthèse de, 3,4-dihydroxyphenylalanine hemoglobine. Biochim.Biophys.Acta 68,496.

Chernavskii, D.W. and Chernavskaya, N.M. (1975) Some theoretic aspects of the problem of life origin. J. theor. Biol. 50,13.

Claus, G. and Nagy, B. (1961) A microbiological examination of some carbonaceous chondrites. Nature 192,594.

Crick, F.H.C. (1963b) The recent excitement in the coding problem. Progr.Nucl.Acid Res.Mol.Biol. 1,163.

Crick, F.H.C. (1966a) The genetic code - yesterday, today and tomorrow. Cold Spring Harbor Symp.Quant.Biol. 31,3.

Crick, F.H.C. (1967a) An error in model building. Nature 213,798.

Crick, F.H.C. (1967b) Origin of the genetic code. Nature 213,119.

Crick, F.H.C. (1968) The origin of the genetic code. J.Mol. Biol. 38,367.

Crick, F.H.C. and Orgel, L.E. (1973) Directed panspermia. Icarus 19,341.

Dayhoff, M.O. (1971) in Chemical Evolution and the Origin of Life (Buvet, R. and Ponnamperuma, C. eds.) Vol. 1, p. 392 Amsterdam:North Holland.

Dunnill, P. (1966) Triplet nucleotide-amino-acid pairing; a stereochemical basis for the division between protein and non-protein amino-acids. Nature 210,1267.

Eigen, M. (1971) Self organization of matter and the evolution of biological macromolecules. Naturwissenschaften 58,465.

Eigen, M. (1973) in The Physicist's Conception of Nature (Mehra, J. ed.) p. 594, Reidel Publishing Co., Dordrecht-Holland.

Epstein, C.J. (1966) Role of the amino-acid "code" and of selection for conformation in the evolution of proteins. Nature 210,25

Evreinova, T.N., Allakhvredov, B.L., Peshenko, V.I. (1974) in The Origin of Life and Evolutionary Biochemistry (Dose, K Fox, S.W., Deborin, G.A., Pavtoskaya, T.E. eds.), p. 89, Plenum Press, New York.

Felsenfeld, G. and Miles, H.T. (1967) The physical and chemica properties of nucleic acids. Ann.Rev.Biochem. 36,407.

Ferris, J.P., Sanchez, R.A. and Orgel, L.E. (1968) Studies in prebiotic synthesis: III. Synthesis of pyrimidines from cyanoacetylene and cyanate. J. Mol. Biol. 33,693.

Fox, S.W. (1964a) The Origin Prebiological Systems and Their Molecular Matrices. (Fox, S.W., ed.), p. 361, Academic Press Inc., New York.

Fox, S.W. (1964b) Thermal polymerisation of amino acids and production of formed microparticles on lava. Nature 201,336.

Fox, S.W. (1968a) A new view of the "synthesis of life". Quart. J. Fla. Acad. Sci. 31,1.

Fox, S.W. (1968b) Spontaneous generation, the origin of life, and self assembly. Curr. Mod. Biol. 2,235.

Fox, S.W. (1969a) a theory of macromolecular and cellular origins. Nature 205, 328.

Fox, S.W. (1969b) Self-ordered polymers and propagative cell-like systems. Naturwissenschaften 36,1.

Fox, S.W. (1974a) Origins of biological information and the genetic code. Mol. Cell. Biochem. 3,129.

Fox, S.W. (1974b) in The Origin of Life and Evolutionary Biochemistry (Dose, K., Fox, S.W., Deborin, G.A., Pavlovskaya, T.E., eds.), Plenum Press, New York/London.

Fox, S.W. (1975) Looking forward to the present. Biosystems 6,165.

Fox, S.W. and Dose, K. eds. (1972) Molecular Evolution and the Origin of Life, W. H. Freman & Co., San Francisco.

Fox, S.W. and Harada, K. (1960a) The thermal copolymerization of amino acids common to protein. J. Am. Chem. Soc. 82,374

Fox, S.W. and Harada, K. (1960b) Thermal copolymerization of amino acids in the presence of phosphoric acid. Arch. Bioc and Biophys. 86,281.

Fox, S.W. and Harada, K. (1961) Synthesis of uracil under conditions of a thermal model of prebiological chemistry. Science 133, 1923.

Fox, S.W. and Krampitz, G. (1964) Catalytic decomposition of glucose in aqueous solution by thermal proteinoids. Nature 203,1362.

Fox, S.W. (1969a) a theory of macromolecular and cellular origins. Nature 205, 328.

Fox, S.W. (1969b) Self-ordered polymers and propagative cell-like systems. Naturwissenschaften 36,1.

Fox, S.W. (1974a) Origins of biological information and the genetic code. Mol. Cell. Biochem. 3,129.

Fox, S.W. (1974b) in The Origin of Life and Evolutionary Biochemistry (Dose, K., Fox, S.W., Deborin, G.A., Pavlovskaya, T.E., eds.), Plenum Press, New York/London.

Fox, S.W. (1975) Looking forward to the present. Biosystems 6,165.

Fox, S.W. and Dose, K. eds. (1972) Molecular Evolution and the Origin of Life, W. H. Freman & Co., San Francisco.

Fox, S.W. and Harada, K. (1960a) The thermal copolymerization of amino acids common to protein. J. Am. Chem. Soc. 82,37

Fox, S.W. and Harada, K. (1960b) Thermal copolymerization of amino acids in the presence of phosphoric acid. Arch. Bio and Biophys. 86,281.

Fox, S.W. and Harada, K. (1961) Synthesis of uracil under conditions of a thermal model of prebiological chemistry. Science 133, 1923.

Fox, S.W. and Krampitz, G. (1964) Catalytic decomposition of glucose in aqueous solution by thermal proteinoids. Nature 203,1362.

Fox, S.W. and Nakashina, T. (1967) Fractionation and characterization of an amidated thermal 1:1:1-proteinoid. Biochim. Biophys.Acta 140,155.

Fox, S.W. and Waehneldt, T. (1968) The thermal synthesis of neutral and basic proteinoids. Biochim. Biophys. Acta 160 , 246.

Fox, S.W. and Wang, C-T. (1968) Melanocyte-stimulating hormone: Activity in thermal polymers of alpha-amino acids. Science 160,547.

Fox, S.W. and Windsor, C.R. (1970) Synthesis of amino acids by the heating of formaldehyde and ammonia. Science 170,984.

Fox, S.W. and Yuyama, S. (1964) Dynamic phenomena in microspheres from thermal proteinoid. Comp.Biochem. & Physiol. 11,317.

Fox, S.W., Harada, K., Woods, K.R. and Windsor, C.R. (1963) Amino acid compositions of proteinoids. Arch. Biochem. & Biophys. 102,439.

Fox, S.W., Harada, K. and Rohlfing, D.L. (1964) in Polyamino Acids, Polypeptides and Proteins, (Stahman, M.A., ed.) p. 47, University of Wisconsin Press, Madison.

Fox, S.W., McCauley, R.J. and Wood, A. (1967) A model of primitive heterotrophic proliferation [molecular evolution]. Comp. Biochem. Physiol. 20,773.

Fox, S.W., Yuki, A., Waehneldt, T.V. and Lacey, J.C. Jr. (1971) in Chemical Evolution and the Origin of Life (Buvet, R. an Ponnamperuma, H.C., eds) Vol. 1,p.252, Amsterdam:North Holland.

Fox, S.W., Hsu, L., Brooke, S., Nakashima, T. and Lacey, J.C. (1972) Experimental models of communications at the molecu lar and microsystemic levels. Int. J. Neurosci.3,183.

Fox, S.W., Harada, K. and Hare, P.E. (1973) Accumulated analysi of amino acid precursors in returned lunar samples. Proc. Earth Lunar Sc. Conf. 2,2241.

Fox, S.W., Jungck, J.R. and Nakshima, T. (1974) From proteinoid microsphore to contemporary cell: Formation of inter- nucleotide and peptide bonds by proteinoid particles. Origins Life 5 , 227.

Gabel, N.W. and Ponnamperuma, C. (1967) Model for origin of monosaccharides. Nature 216,453.

Gauss, D.H., van der Haar, F., Maekicke, A. and Cramer, F. (197 Recent results of tRNA research. Ann. Rev. Biochem. 40,10

Glansdorff, P. and Prigogine,I. (1971) Thermodynamic Theory of Structure, Stability and Fluctuations. Wiley-Interscience New York.

Goel, N. and Islam, S. (1977) Journal of Theoretical Biology (in press).

Goel, N.S., Yčas, M. (1975) The error catastrophe hypothesis with reference to ageing and the evolution of the protein synthesizing machinery. J. theor.Biol. 55, 245.

Goldberg, A.L. and Wilks, R.E. (1966) Genetic Code: Aspects o organization. Science 153,420.

Handschuh, G.J. and Orgel, L.E. (1973) Struvite and prebiotic phosphorilation. Science 179, 483.

Harada, K. and Fox, S.W. (1964a) in The Origins of Prebiologic Systems and their Molecular Matrices, p.187, Academic Press Inc., New York.

Harada, K. and Fox, S.W. (1964b) in The Origins of Prebiologic Systems and their Molecular Matrices, p. 289, Academic Press Inc., New York.

Harpold, M.A. and Calvin, M. (1968) AMP on an insoluble solid support. Nature 219, 486.

Harpold, M.A. and Calvin, M. (1973) Amino acid-nucleotide inte actions on an insoluble solid support: I. A simple model the amino acid acceptor terminus of a tRNA. Biochim. Bio Acta 308 , 117.

Hartman, H. (1975) Speculations on the evolution of the genetic code. Origins Life 6, 423.

Hayakawa, T., Windsor, C.R. and Fox, S.W. (1967) Copolymerisa tion of the lenchs anhydrides of the eighteen amino acids common to protein. Arch. Biochem. Biophys. 118(2), 265.

Herbig, G.H. (1974) Interstellar smog. Amer. Scientist 62, 20

Hinegardner, R.T. and Engelberg, J. (1963) Rationale for a universal genetic code. Science 142, 1083.

Hinegardner, R.T. and Engelberg, J. (1964) Comment on a criticism by Woese (1964). Science 144, 1031.

Hoffmann, G.W. (1974) On the origin of the genetic code and the stability of the translation apparatus. J. Mol. Biol. 68, 349.

Hoffmann, G.W. (1975) The stochastic theory of the origin of the genetic code. Annual Reviews of Physical Chemistry 26, 123.

Hoffmann, G.W. and Pörschke, D. (1973a) Cooperative nonenzymic base recognition. Thermodynamics of the helix-coil transition of a monomer-polymer double helix. Biopolymers 12, 1611.

Hoffmann, G.W. and Pörschke, D. (1973b) Cooperative nonenzymic base recognition. Kinetics of the binding of a base monome to a complementary polynucleotide template. Biopolymers 12, 1625.

Holmquist, W.R. (1975) Directions from compositional randomness in eukaryotic and prokaryotic proteins: the hypothesis of selective-stochastic stability and a principle of charge conservation. J. Mol. Evol. 4, 277.

Horowitz, N.H. and Miller, S.L. (1962) Current theories on the origin of life. Fortschr. Chem. Org. Naturstoffe 20, 423.

Howard, F.B., Frazier, J., Lipsitt, M.N. and Moles, H.T. (1964) Infrared demonstration of two- and three-strand helix formation between poly C and guanosine mononucleotides and oligonucleotides. Biochem. Biophys. Res. Comm. 17, 93

Howard, F.B., Frazier, J., Singer, M.F. and Miles, H.T. (1966) Helix formation between polyribonucleotides and purines, purine nucleosides and nucleotides. II. J. Mol. Biol. 16, 415.

Hsu, L.L., Brooke, S. and Fox, S.W. (1971) Conjugation of proteinoid microsphores: A model of primordal communication. Curr. Mod. Biol. 4 , 12.

Huang, W.M. and Tso, P.O.P. (1966) Physico-chemical basis of the recognition process in nucleic acid interaction. I. Interaction of polyuridylic acid and nucleosides. J. Mol. Biol. 16, 523.

Ibanez, J.D., Kiball, A.P. and Oro, J. (1971) Possible prebiotic condensation of mononucleotides by cyanamide. Science 173 , 444.

Ito, Y. and Bowman, R.L. (1971) Long-wavelength photoproduction of amino acids on the primitive earth. Science 173 , 417.

Jong, B.H.G.de (1949) in Colloid Science (Van Niel, C.B., ed.), p.423, Elsevier, New York, Vol. 2.

Jukes, T.H. (1965) Coding triplets and their possible evolutionary implications. Biochem. Biophys. Res. Commun. 19, 391.

Jukes, T.H. (1966) Molecules and Evolution, Columbia University Press, New York

Jukes, T.H. (1969a) Recent advances in studies of evolutionary relationships between proteins and nucleic acids. Space Life Sci. 1 , 469.

Jukes, T.H. (1969b) Recent problems in the genetic code. Curre Topics in Microbiology 49, 178.

Jukes, T.H. (1973a) Possibilities for the evolution of the gene tic code from a preceding form. Nature 246 , 22.

Jukes, T.H. (1973b) Arginine as an evolutionary intruder into protein synthesis. Biochem. Biophys. Res. Commun. 53, 709

Jukes, T.H. (1974) On the possible origin and evolution of the genetic code. Origins Life 5 , 331.

Jukes, T.H. and Holmquist, R. (1972) Evolution of transfer RNA molecules as a repetitive process. Biochem. Biophys. Res. Commun. 49, 212.

Jukes, T.H., Holmquist, R., Moise, H. (1975) Amino acid composition or proteins: Selection against the genetic code. Science 189, 50.

Jungck, J.R. (1971) Pre-Darwinian and non-Darwinian evolution of proteins. Curr. Mod. Biol. 3 , 307.

Kaplan, R.W. (1972) Ursprung des Lebens durch Zufall. Umschau 14, 456.

Kenyon, D.H. and Steinman, G. (1969) Biochemical Predestination McGraw-Hill, New York.

Kenyon, D.H. (1973) A theory of biogenesis. Science 179, 789.

Kim, S.H., Quigley, G.J., Suddath, F.L., McPherson, A., Sneden, D., Kim, J.J., Weinzierl, J. and Rich, A. (1973) Science 179, 285

Kimball, A.P. and Oro, J. (eds.) (1971) Prebiotic and Biochemical Evolution, North Holland Publ. Co., Amsterdam.

King, J.L. and Jukes, T.W. (1969) Non-Darwinian evolution. Science 164, 788.

Kirkwood, T.B.L., Holliday, R. (1975) The stability of the translation apparatus. J. Molec. Biol.

Kondo, N.S., Holmes, H.M., Stempel, L.M., Tso, P.O.P. (1970) Influence of the phosphodiester linkage (3'-5', 2'-5' and 5'-5') on the conformation of dinucleoside monophosphate. Biochemistry 9, 3479.

Krampitz, G. and Fox, S.W. (1969) The condensation of the adenylates of the amino acids common to protein. Proc. Nat. Acad. Sci. U.S.A. 62, 399.

Kvenvolden, K.A. (1974) in The Origin of Life and Evolutionary Biochemistry (Dose, K., Fox, S.W., Deborin, G. and Pavlovskaya, T.W., eds.) p.301, Plenum Press, New York.

Kvenvolden, K., Lawless, J., Pering, K., Peterson, E., Flores, J Ponnamperuma, C., Kaplan, I.R. and Moore, C. (1970) Evidence for extraterrestrial a.a. and hydrocarbons in Murchison meteorite. Nature 228 , 923.

Kvenvolden, K., Lawless, J.G. and Ponnamperuma, C. (1971) Nonprotein amino acids in the Murchison meteorite. Proc. Nat. Acad. Sci. 68, 486.

Lacey, J.C. Jr., and Pruitt, K.M. (1969) Origin of the genetic code. Nature 223, 799.

Lacey, J.C., Weber, A.L., White, W.E. (1975) A model for the c evolution of the genetic code and the process of protein synthesis: Review and assessment. Origins Life 6 , 27

Latt, S.A. and Sober, H.A. (1967a) Protein-nucleic acid interactions. II. Oligopeptide polyribonucleotide binding studies. Biochemistry 6, 3293.

Latt, S.A. and Sober, H.A. (1967b) Protein-nucleic acid interac III. Cation effect on binding strength and specificity. Biochemistry 6, 3307.

Lawless, J.G., Kvenvolden, K.A., Peterson, E., Ponnamperuma, C. and Moore, C. (1971) Amino acids indigenous in the Murray meteorite. Science 173 , 626.

Lehninger, A.L. (1970) Biochemistry, Worth and Co., New York.

Long, M. and Felsenfeld, G. (1966) The preferential interaction of polylysine and polyarginine with specific base sequence in DNA. Proc. Natl. Acad. Sci. U.S.A. 56, 1325.

Lipman, F. (1971) Attempts to map a process evolution of peptid biosynthesis. Science 173, 875.

Lohrmann, R. and Orgel, L.E. (1968) Prebiotic synthesis: Phosphorylation in aqueous solution [biochemical evolution]. Science 161 , 64.

Lohrmann, R. and Orgel, L.E. (1971) Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. Science 171 , 490.

Mackay, A.L. (1967) Optimization of the genetic code. Nature 216, 159.

Marshak, A. (1967) The harmonic oscillator theory applied to the code translation problem. J. Theor. Biol. 17 , 12.

Melcher, G. (1970) A new hypothesis on the evolution of the genetic code. Biophysik 7 , 25.

Melcher, G. (1974) Stereospecificity of the genetic code. J. M Evol. 3, 121.

Meuller, G. (1972) Organic microsphores from the precambrian of South-West Africa. Nature 235 , 90.

Miller, S.L. (1953) Production of amino acids under possible primitive earth conditions. Science, 117, 528.

Miller, S.L. (1955) Production of some organic compounds under possible primitive earth conditions. J. Amer. Chem. Soc. 77, 2351.

Miller, S.L. (1957a) The mechanism of synthesis of amino acids by electric discharges. Biochim. Biophys. Acta 23, 480.

Miller, S.L. (1957b) The formation of organic compounds on the primitive earth. Ann. N.Y. Acad. Sci. 69, 260.

Miller, S.L. (1959) The Origin of Life on the Earth (Oparin, A.I ed.), p. 123, Pergamon, Oxford.

Mikelsaar, H.N. (1975) A concept of amino acid archaeorelation: Origin of life and the genetic code. J. Theor. Biol. 50, 2

Miquel, J., Brooke, S. and Fox, S.W. (1971) Assembly of micro-sphores from acidic proteinoids and histones or histone-lik proteinoids. Curr. Mod. Biol. 3 , 299.

Modzeleski, V.E., Modzeleski, J.E., Mohammad, M.A.J., Nagy, L.A., Nagy, B., McEwan, W.S., Urey, H. and Hamilton, P.B. (1973) Carbon compounds in pyrolysates and amino acids in extracts of Apollo 14 lunar samples. Nature Physical Scien 242, 50.

Moravek, J., Kopecky, J. and Skoda, J. (1968a). Collection Czech. Chem. Commun. 33, 960. Cited by Orò (1974).

Moravek, J., Kopecky, J. and Skoda, J. (1968b). Collection Czech. Chem. Commun. 33, 4407. Cited by Orò and Stephen-Sherwood (1974).

Morrison, P. (1962) Carbonaceous "Snowflakes" and the origin of Life. Science 135, 663.

Nagyvary, J. and Fendler, J.H. (1974) Origin of the genetic code A physical-chemical model of primitive codon assignments. Origins Life 5, 357.

Nakashima, T. and Fox, S. (1972) Selective condensation of amino-acyl adenylates by nucleotproteinoid microparticles. Proc. Nat. Acad. Sci. U.S.A. 65, 106.

Novák, V. (1974) in The Origin of Life and Evolutionary Biochemist (Dose, K., Fox, S.W., Debovin, G.A. and Pavlovshaya, T.E., eds.) p. 355, Plenum Press, New York.

Novák, V. and Liebl, V. (1975) On the question of the origin and evolution of the genetic system. Origins Life 6, 269.

Ohba, Y. (1966) Structure of nucleohistone. II. Thermal denaturation. Biochim. Biophys. Acta 123, 84.

Oparin, A.I. (1957) , The Origin of Life on Earth 3rd ed. , Oliver & Boyd, Edinburg.

Oparin, A.I. (1965) The origin of life and the origin of enzyme Advances in Enzymology 27, 347.

Oparin, A.I. (1966) A study of the enzymatic processes in coacervation systems. Rev. Roumaine Biochim. 3, 111.

Orgel, L.E. (1963) The maintenance of the accuracy of protein synthesis and its relevance to ageing. Proc. Nat. Acad. Sci. U.S.A. 49, 517.

Orgel, L.E. (1968) Evolution of the genetic apparatus. J. Mol. Biol. 38, 381.

Orgel, L.E. (1972) A possible step in the origin of the genetic code. ISR J. Chem. 10, 287.

Orgel, L.E. (1973) Aging of clones of mammalian cells. Nature 243, 441.

Orgel, L.E. and Sulston, J.E. (1971) in Prebiotic and Biochemical Evolution (Kimball, A.P. and Oro, J., eds.) p. 89, North Holland Publ. Co., Amsterdam.

Oró, J. (1960) Symthesis of adenine from ammonium cyanide. Biochem. Biophys. Res. Commun. 2, 407.

Oró, J. (1961) Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. Nature 191, 1193.

Oró, J. (1963) Studies in experimental organic cosmochemistry. Ann. N.Y. Acad. Sci. 108, 464.

Oró, J., Kimball, A.P. (1961) Synthesis of purines under possib primitive Earth conditions. I. Adenine from hydrogen cyanide. Arch. Biochem. Biophys. 94, 217.

Oró, J., Kimball, A.P. (1962) Synthesis of purines under possib primitive Earth conditions. II. Purine intermediates from hydrogen cyanide. Arch. Biochem. Biophys. 96, 293.

Oró, J. and Stephen-Sherwood, E. (1974) The prebiotic synthesis of oligonucleotides. Origins Life 5, 159.

Osterberg, R., Orgel, L.E. (1973) Polyphosphate and trimeta- phosphate formation under potentially prebiotic conditions J. Mol. Evol. 1, 241.

Paecht-Horowitz, M. (1974) The possible role of clays in pre- biotic peptide synthesis. Origins Life 5, 173.

Paecht-Horowitz, M., Berger, J. and Katchalsky, A. (1970) Prebiotic synthesis of polypeptides in heterogeneous polycondensation of amino-acid adenylates. Nature 228, 636

Papentin, F. (1973) A Darwinian evolutionary system: II. Experi- ments on protein evolution and evolutionary aspects of the genetic code. J. Theor. Biol. 39, 417.

Park, W.K., Hochitim, A.R. and Ponnamperuma, C. (1975) Organic synthesis by quench reactions. Origins Life 6, 99.

Pattee, H.H. (1961) On the origin of macromolecular sequences. Biophys. Jour. 1, 683.

Pattee, H.H. (1967) Quantum mechanics, heredity and the origin of life. J. Theor. Biol. 17, 410.

Pattee, H.H. (1967) in Towards a Theoretical Biology. (Prolegomena, I., Waddington, C.H., eds.) p. 67, Edinburgh University Press and Aldine Publ. Co., Chicago.

Pattee, H.H. (1971) Physical theories of biological co-ordinati Quart. Rev. of Biophys. 4, 255.

Pattee, H.H. (1972) Physical problems of decision-making constraints. Int. J. Neurosci. 3, 99.

Pavlovskaya, T.E., Telegina, T.A., Sokol'skaya, A.V. and El'pin J.E. (1971) [Amino acid formation under the effect of UV light on a formaldehyde and ammonium nitrate fog] IZV Akad Nauk. SSSR Ser. Biol. 6, 922

Polanyi, M. (1967) Life transcending physics and chemistry. Chem. Eng. News, 45, 55.

Pelc, S.R. (1965) Correlation between coding-triplets and amino-acids. Nature 207, 597.

Pelc, S.R., Welton, M.G.E. (1966) Stereochemical relationships between coding triplets and amino-acids. Nature 209, 868.

Pongs, O. and Ts'o, P.O.P. (1971) Polymerisation of unprotected 2'-deoxyribonucleoside 5'-phosphates at elevated temperature. J. Am. Chem. Soc. 93, 5241.

Ponnamperuma, C. (1965) in Origins of Prebiological Systems and of their Molecular Matrices. (Fox, S.W., ed.) p. 221 Academic Press, N.Y.

Ponnamperuma, C. and Kirk, P. (1964) Synthesis of deoxyadenosine under simulated primitive earth conditions. Nature 203, 400.

Ponnamperuma, C. and Mack. R. (1965) Nucleotide synthesis under possible primitive earth conditions. Science 148, 1221.

Ponnamperuma, C. and Mariner, R. (1963) The formation of ribose and deoxyribose by ultraviolet irradiation of formaldehyde in water. In: 11th Annual meeting of the Radiation Resear Society. Radiation Res. 19, 183.

Ponnamperuma, C. and Peterson, E. (1965) Peptide synthesis from amino acids in aqueous solution. Science 147, 1572.

Ponnamperuma, C.A., Lemmon, R.M. and Calvin, M. (1962) Chemical effect of ionizing radiation on cytosine. Science 137, 605

Ponnamperuma, C., Mariner, R. and Sagan, C. (1963a) Formation of adenosine by ultraviolet irradiation of a solution of adenine and ribose. Nature 198, 1199.

Ponnamperuma, C., Sagan, C. and Mariner, R. (1963b) Synthesis of adenosine triphosphate under possible primitive earth conditions. Nature 199, 222.

Ponnamperuma, C., Lemmon, R.M., Mariner, R. and Calvin, M. (196 Formation of adenine by electron irradiation of methane, a nia and water. Proc. Nat. Acad. Sci. USA 49, 737.

Ponnamperuma, C., Young, R.S., Munoz, E.F. and McCaw, B.V. (196 Guanine: Formation during the thermal polymerization of amino acids. Science 143, 1449.

Pörschke, D. and Eggers, F. (1972) Thermodynamics and kinetics of base-stacking interactions. Eur. J. Biochem. 26, 490.

Pörschke, D. and Eigen, M. (1971) Co-operative non-enzymic base recognition: III. Kinetics of the helix-coil transition of the oligoribouridylic-oligoriboadenylic acid system and of oligo-riboadenylic acid alone at acidic pH. J. Mol. Bi 62, 361.

Pörschke, D., Hoffman, G.W. and Senear, A. (1973) Double helical complex formed from a polynucleotide and a complementary monomer. Nature NB 242, 45.

Raska, M., Mandel, M. (1972) Is there a physical chemical basis for the present genetic code? J. Mol. Evol. 2, 38.

Reid, C. and Orgel, L.E. (1967) Synthesis of sugars in potential prebiotic conditions. Nature 216, 455.

Renz, M., Lohrmann, R. and Orgel, L.E. (1971) Catalysts for the polymerization of adenosine cyclic 2', 3'-phosphate on a poly (U) template. Biochim. Biophys. Acta 240, 463.

Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C. and Klug, A. (1974) Structure of yeast phenylalamine tRNA at 3 Å resolution. Nature 250, 546.

Rohlfing, D.L. (1970) Catalytic activities of thermally prepared poly-α-amino acids: Effect of aging. Science 169, 998.

Rohlfing, D. (1975) Coacervate-like microsphores from lysine-rich proteinoid. Origins Life 6, 203.

Rohlfing, D.L. and Fox, S.W. (1967) The catalytic activity of thermal polyanhydro-α-amino acids for the hydrolysis of p-nitrophenyl acetate: Catalysis by thermal polyamino acids. Arch. Biochem. Biophys. 118, 122.

Rohlfing, D.L. and Fox, S.W. (1967) The inactivation of catalytically active thermal polyanhydro-α-amino acids. Arch. Biochem. Biophys. 118, 127.

Ryan, J.W. and Fox, S.W. (1973) Activation of glycine by ATP, a divalent cation, and proteinoid microsphores. Biosystem 5, 115.

Reanney, D.C. and Ralph, R.K. (1967) A speculation on the origi of the genetic code. J. Theor. Biol. 15, 41.

Samuel, E. (1972) Order in Life , Prentice-Hall, Englewood Cliffs, N.J.

Salthe, S.N. (1972) Evolutionary Biology, Holt, Rinehart and Winston, New York.

Sanchez, R.A. and Orgel, L.E. (1970) Studies in prebiotic synthesis: V. Synthesis and photoanomerization of pyrimidine nucleosides. J. Mol. Biol. 47, 531.

Sanchez, R., Ferris, J. and Orgel, L.E. (1966a) Conditions for purine synthesis: Did prebiotic synthesis occur at low temperatures? Science 153, 72.

Sanchez, R.A., Ferris, J.P. and Orgel, L.E. (1966b) Cyanoacetyl in prebiotic synthesis. Science 154, 784.

Sanchez, R.A., Ferris, J.P. and Orgel, L.E. (1967) Studies in prebiotic synthesis: II. Synthesis of purine precursors an amino acids from aqueous hydrogen cyanide. J. Mol. Biol. 30, 223.

Sanchez, R.A., Ferris, J.P. and Orgel, L.E. (1968) Studies on prebiotic synthesis: IV. Conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. J. Mol. Biol. 38, 1

Saxinger, C, Ponnamperuma, C. (1971) Experimental investigation on the origin of the genetic code. J. Mol. Evol. 1, 63.

Saxinger, C. and Ponnamperuma, C. (1974) Interactions between amino acids and nucleotides in the prebiotic milieu. Origins Life 5, 189.

Saxinger, C., Ponnamperuma, C. and Woese, C. (1971) Evidence f the interaction of nucleotides with immobilized amino aci and its significance for the origin of the genetic code. Nature NB 234, 172.

Schaap, T. (1971) Dual information in DNA and the evolution of the genetic code. J. Theor. Biol. 32, 293.

Schneider-Bernloehr, H., Lohrmann, R., Sulston, J., Weimann, B Orgel, L.E. and Miles, H.T. (1968) Non-enzymic synthesis deoxyadenylate oligonucleotides on a poly u ridylate templa J. Mol. Biol. 37, 151.

Schutzenberger, M-P, Gavaudan, P. and Besson, J. (1969) [The ex istence of a certain correlation between the molecular weight of amino acids and the number of triplets coding for them]. C R Hebd Seances Acad. Sci. SER D. Sci. Natur (Paris) 268, 1342.

Schwartz, A.W. (1972) Prebiotic phosphorylation-nucleotide synthesis with apatite. Biochim. Biophys. Acta 281, 477.

Schwartz, A.W. and Fox, S.W. (1964) Thermal synthesis of internucleotide phosphodiester linkages. Biochim. Biophys. Acta 87, 694.

Schwartz, A.W. and Fox, S.W. (1967) Condensation of cytidylic acid in the presence of polyphosphoric acid. Biochim. Biophys. Acta 134, 9.

Schwartz, A.W., Bradley, E. and Fox, S.W. (1964) in The Origins of Prebiological Systems and of Their Molecular Matrices. p. 317, Academic Press, New York.

Shook, L.K. and Rohlfing, D.L. (1972) The catalytic hydrolysis of p-nitrophenyl acetate by thermally prepared copolymers of lysine and histidine. Curr. Mod. Biol. 5, 43.

Smith, T.F. (1969) The genetic code, information density, and evolution. Math. Biosci. 4, 179.

Sobor, H.A., Schlossman, S.F., Yaron, A., Latt, S.A. and Rushizky, G.W. (1966) Protein-nucleic acid interaction. I. Nuclease-resistant polylysine-ribonucleic acid complexe Biochemistry 5, 3608.

Sonneborn, T.M. (1965) in Evolving Genes and Proteins. (Bryson, V. and Vogel, H.J. eds.) p. 377, Academic Press, New York.

Spitnik, P., Lipshitz, R. and Chargaff, E. (1955) Studies on nucleoproteins. III. Deoxyribonucleic acid complexes with basic polyelectrolytes and their fractional extraction. J. Biol. Chem. 215, 765.

Steinman, G. and Cole, M.N. (1967) Synthesis of biologically pertinent peptides under possible primordial conditions. Proc. Nat. Acad. Sci. USA 58, 735.

Steinman, G., Lemmon, R.M. and Calvin, M. (1964) Cyanamide: A possible key compound in chemical evolution. Proc. Natl. Acad. Sci. 52, 27.

Steinman, G., Lemmon, R.M. and Calvin, M. (1965) Dicyandiamide: Possible role in peptide synthesis during chemical evolution. Science 147, 1574.

Steinman, G., Kenyon, D.H. and Calvin, M. (1966) The mechanism and protobiochemical relevance of dicyanamide-mediated peptide synthesis. Biochim. Biophys. Acta 124, 339.

Stephen-Sherwood, E., Oró, J. and Kimball, A.P. (1971) Thymine: A possible prebiotic synthesis. Science 173, 446.

Subbaraman, A.S., Kazi, Z.A. and Choughuley, A.S.V. (1972) Reactions of aldehydes with glycine as possible prebiotic events. Indian J. Biochem. Biophys. 9, 268.

Sulston, J., Lohrmann, R., Orgel, L.E. and Miles, H.T. (1968a) Non enzymatic synthesis of oligoadenylate on a polyuridylic acid template. Proc. Nat. Acad. Sci. USA 59, 726.

Sulston, J., Lohrmann, R., Orgel, L.E. and Miles, H.T. (1968b) Specificity of oligonucleotides synthesis directed by polyuridylic acid. Proc. Nat. Acad. Sci. USA 60, 409.

Sulston, J., Lohrmann, R., Orgel, L.E., Schneider-Bernloehr, H. and Weimann, B.J. (1969) Non-enzymic oligonucleotide-synthesis on a polycytidylate template. J. Mol. Biol. 40, 227.

Tapiero, C.M. and Nagyvary, J. (1971) Prebiotic formation of cytidine nucleotides. Nature 231, 42.

Urey, H.C. (1966) Biological material in meteorites. Science 151, 157.

Volkenstein, M.V. (1966) The genetic coding of protein structure. Biochim. Biophys. Acta 119, 421.

Volkenstein, M.V. (1967). Physics of Enzymes, Nauka, Moscow.

Waehneldt, T.V. and Fox, S.W. (1967) Phosphorylation of nucleosides with polyphosphoric acid. Biochim. Biophys. Acta 134, 1.

Waehneldt, T.V. and Fox, S.W. (1968) The binding of basic proteinoids with organismic or thermally synthesized polynucleotides. Biochim. Biophys. Acta 160, 239.

Wagner, K.G., Arav, R. (1968) On the interaction of nucleotides with poly-L-lysine and poly-L-arginine. The influence of the nucleotide base on binding behaviour. Biochemistry 7, 1771.

Walker, G.W.R. (1974) Genetics and the origin of the genetic code. Origins Life 5, 351.

Watson, J.D. (1970) Molecular Biology of the Gene, 2nd ed. Benjamin, Menlo Park, Calif.

Weimann, B.J., Lohrmann, R., Orgel, L.E., Bernloehr-Schneider, H. and Sulston, J.F. (1968) Template-directed synthesis with adenosine-5'-phosphoimidzaolide. Science 161, 387.

Welton, M.G.E., Pelc, S.R. (1966) Specificity of the stereochemical relationship between ribonucleic acid-triplets and amino acids. Nature 209, 870.

Wigner, E.P. (1961) in The Logic of Personal Knowledge (Shils, E., ed.) p.200, The Free Press, Glencoe, Ill.

Woese, C.R. (1964) Universality of the genetic code. Science 144, 1030.

Woese, C.R. (1965a) Order in the genetic code. Proc. Natl. Acad Sci. USA 54, 71.

Woese, C.R. (1965b) On the evolution of the genetic code. Proc Natl. Acad. Sci. 54, 1546.

Woese, C.R. (1967) The Genetic Code. Harper & Rowe, New Yor

Woese, C.R. (1968) The fundamental nature of the genetic code: Prebiotic interactions between polynucleotides and polyamino acids or their derivatives. Proc. Natl. Acad. Sci. USA 59, 110.

Woese, C.R. (1970a) The problem of evolving a genetic code. Bioscience 20, 471.

Woese, C.R. (1970b) Molecular mechanism of translation: A reciprocating ratchet mechanism. Nature 226, 817.

Woese, C.R. (1971) Evolution of macromolecular complexity. J. Theor. Biol. 33, 29.

Woese, C.R. (1973a) Evolution of nucleic acid replication: The possible role of simple repeating sequence polypeptides therein. J. Molec. Evol. 2, 205.

Woese, C.R. (1973b) The rotating ribosome: A gross mechanical model for translation. J. Theor. Biol. 38, 203.

Woese, C.R., Dugre, D.H., Dugre, S.A., Kondo, M. and Saxinger, W (1966a) On the fundamental nature and evolution of the genetic code. Cold Spring Harbor Symp. Quant. Biol. 31, 72

Woese, C.R., Dugre, D.H., Saxinger, W.C. and Dugre, S.A. (1966b) The molecular basis for the genetic code. Proc. Natl. Acad. Sci. USA 55, 966.

Wong, J.T-F (1975) A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. USA 72, 1903.

Yčas, M. (1955) A note on the origin of life. Proc. Natl. Acad. Sci. USA 45, 1721.

Yčas, M. (1969) The Biological Code, North-Holland, Amsterdam.

Yčas, M. (1974) On earlier states of the biochemical system. J. Theor. Biol. 44, 145.

Young, R.S. and Ponnamperuma, C. (1964) Life: origin and evolutio Science 143, 384.

Yuky, A. and Fox, S.W. (1969) Selective formation of particles by binding of pyrimidine polyribonucleotides or purine polyribonucleotides with lysine-rich or arginine-rich proteinoids. Biochem. Biophys. Res. Commun. 36, 657.