Die Naturwissenschaften 65. Jahrgang Heft 7 Juli 1978

The Hypercycle

A Principle of Natural Self-Organization

Part C: The Realistic Hypercycle

Manfred Eigen
Max-Planck-Institut für biophysikalische Chemie, D-3400 Göttingen

Peter Schuster
Institut für theoretische Chemie und Strahlenchemie der Universität, A-1090 Wien

The proposed model for a 'realistic hypercycle' is closely associated with the molecular organization of a primitive replication and translation apparatus. Hypercyclic organization offers selective stabilization and evolutive adaptation for all geno- and phenotypic constituents of the functionally linked ensemble. It originates in a molecular quasi-species and evolves by way of mutation and gene duplication to greater complexity. Its early structure appears to be reflected in: the assignment of codons to amino acids, in sequence homologies of tRNAs, in dual enzymic functions of replication and translation, and in the structural and functional organization of the genome of the prokaryotic cell.

XI. How to Start Translation?

"The origin of protein synthesis is a notoriously difficult problem. We do not mean by this the formation of random polypeptides but the origin of the synthesis of polypeptides directed, however crudely, by a nucleic acid template and of such a nature that it could evolve by steps into the present genetic code, the expression of which now requires the elaborate machinery of activating enzymes, transfer RNAs, ribosomes, factors, etc."
Our subject could not be characterized more aptly than by these introductory phrases, quoted from a recent paper by F.H.C. Crick, S. Brenner, A. Klug and G. Pieczenik [3].

Let us for the time being assume that a crude replication and translation machinery, functioning with adequate precision, and adapted to a sufficiently rich alphabet of molecular symbols, has come into existence by some process not further specified, e.g., by self-organization or creation, in Nature or in the laboratory. Let us further suppose an environment which supplies all the activated, energy-rich material required for the synthesis of macromolecules such as nucleic acids and proteins, allowing both reproduction and translation to be spontaneous processes, i.e., driven by positive affinities. Would such an ensemble, however it came into existence, continue to evolve as a Darwinian system? In other words, would the system preserve indefinitely the information which it was given initially and improve it further until it reaches maximal functional efficiency?
In order to apply this question to a more concrete situation let us consider the model depicted in Figure 45. The plus strands of a given set of RNA molecules contain the information for a corresponding number of protein molecules. The products of translation can fulfill at least the following functions: (1) One protein acts as an RNA-polymerase similar to the specific replicases associated with various RNA phages. Its recognition site is adapted to a specific sequence or structure occurring in all plus and minus strands of the RNAs; in other words, it reproduces efficiently only those RNA molecules which identify themselves as members of the particular ensemble. (2) The other translation products function as activating enzymes, which assign and link various amino acids uniquely to their respective RNA adaptors, each of which carries a defined anticodon. The number of different

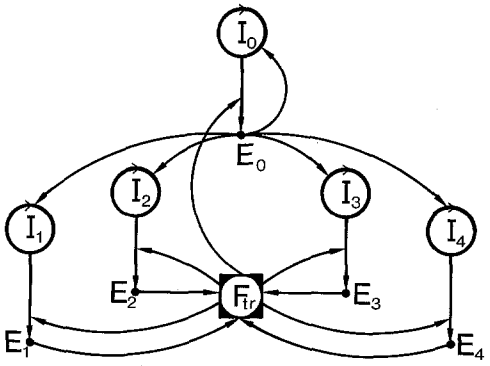


Fig. 45. A minimum model of primitive translation involves a messenger I_0 encoding a replicase E_0, which is adapted to recognize specifically the sequences I_0 to I_4. The plus strands of I_1 to I_4 encode four synthetase functions E_1 to E_4, while the minusstrands may represent the adapters (tRNAs) for four amino acids. Such a system, although it includes all functions required for translation and self-reproduction, is unstable due to internal competition. Coherent evolution is not possible, unless I_0 to I_4 are stabilized by a hypercyclic link

amino acids and hence of adaptors is adjusted to match the variety of codons appearing in the messenger sequences, i.e., the plus strands of the RNAs, so as to yield a 'closed' translation system with a defined code. It does not necessarily comprise the complete genetic code, as it is known today, but rather may be confined to a – functionally sufficient – smaller number of amino acids (e.g., four), utilizing certain constraints on the codon structure in order to guarantee an unambiguous read-off. The adaptors may be represented by the minus strands of the RNA constituents, or, if this should be too restrictive a condition, they could be provided along with further machinery, such as ribosomes, in the form of constant environmental factors similar to the host factors assisting phage replication and translation inside the bacterial cell.

At first glance, we might find comfort in the thought that the system depicted in Figure 45 appears to be highly functionally interwoven; all I_i are supported catalytically by the replicase E_0, which in turn owes its existence to the joint function F_t of the translation enzymes E_1 to E_4 without which it could not be translated from I_0. The enzymes E_1 to E_4, of course, utilize this translation function for their own production too, but being the translation products of I_1 to I_4, they are finally dependent also upon E_0 or I_0 respectively.

However, a detailed analysis shows that the couplings present are not sufficient to guarantee a mutual stabilization of the different genotypic constituents I_i. The general replicase function exerted by E_0 and the general translation function F_{tr} are represented in all differential equations by the same term. The equations then reduce to those for uncoupled competitors, multiplied by a common time function $f(t)$. The system, which initially functions quite well, is predestined to deteriorate, owing to internal competition. A typical set of solution curves, obtained by numerical integration of the rate equations, is shown in Figure 46.

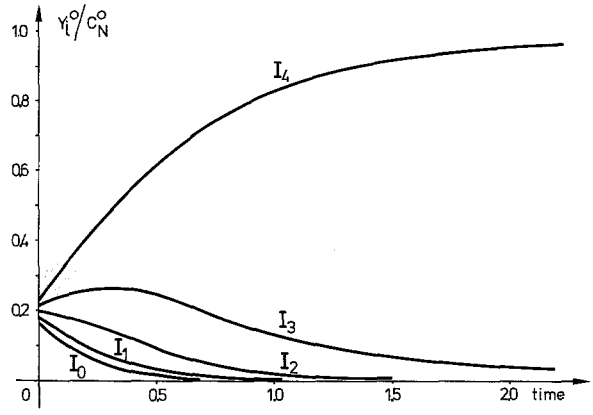


Fig. 46. Solution curves for a system of differential equations simulating the model represented in Figure 45. In this particular example, it is assumed that initial concentrations and autocatalytic-reproduction-rate constants increase linearly from I_0 to I_4, while the other parameters – such as translation-rate constants ($I_i \xrightarrow{k_i} E_i$), amino acid assignments (contribution of E_1, E_2, E_3 E_4 to F_{tr}) or enzyme-substrate-complex stabilities ($I_i + E_0 \overset{K_i}{\rightleftharpoons} I_i \cdot E_0$), etc. – are identical for all reaction partners. The time course of the relative population numbers (y_i^0/c_N^0) reflects the competitive behavior. The most efficiently growing template (I_4) will supersede all others and finally dominate ($y_4^0/c_N^0 \to 1$). However, since both replication (represented by E_0) and translation function (contributions of E_1, E_2 and E_3 to F_{tr}) disappear, I_4 will also die out. The total population is bound to deteriorate ($c_N^0 \to 0$)

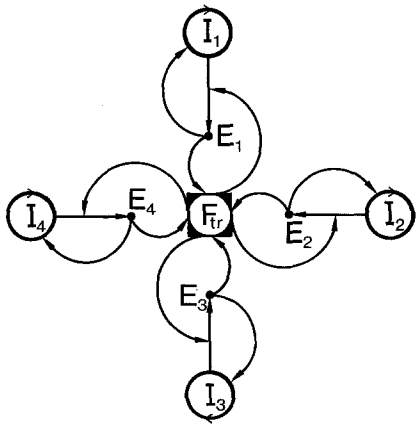


Fig. 47. In this alternative model for primitive replication and translation, the enzymes E_1 to E_4 are assumed to have dual functions, i.e., as specific replicases of their own messengers and as synthetases for four amino acid assignments. The fate of the system is the same as that of the system depicted in Figure 45, since the messengers are highly competitive

Another example of this kind is represented in Figure 47. Here all messengers produce their own specific replicases E_1 to E_4, which also provide synthetase functions (F_{tr}). Again, this coupling by means of a concomitant translation function does not suffice to stabilize the ensemble. The answer to our question, whether the mere presence of a system of messengers for replicase and translation functions and of translation products is sufficient for its continuous existence and evolution, is that unless a particular kind of coupling among the different replicative constituents I_i is introduced, such systems are not stable, despite the fact that they contain all required properties for replication and translation. Even if all partners were selectively equivalent (or nearly equivalent) and hence were to coexist for some time (depending on their population size), they could not evolve in a mutually controlled fashion and hence would never be able to optimize their functional interaction. Their final fate would always be deterioration, since an occasional selective equivalence cannot be coherently maintained over longer periods of evolution unless it is reinforced by particular couplings.

Knowing the results of part B, we are, of course, not surprised by this answer. A closer inspection of the particular linkages provided by the functions of replication and translation enzymes does not reveal any hypercyclic nature. Therefore these links cannot establish the mutual control of population numbers that is required for the interrelated evolution of members of an organized system. The couplings present in the two systems studied can be reduced to two common functions, which, like environmental factors, influence all partners in exactly the same way and hence do not offer any possibility of mutual control.

The above examples are typical of what we intend to demonstrate in this article, namely, that

1. In the early phases of evolution, characterized by low fidelities of replication and translation as well as by the initially low abundance of efficiently replicating units, hypercyclic organization offers large relative advantages over any other kind of (structural) organization (Sect. XV), and

2. That hypercyclic models *can* indeed be built to provide realistic precursors of the reproduction and translation apparatus found in present prokaryotic cells (Sect. XVI).

How could we envisage an origin of translation, given the possible existence of reproducible RNA molecules as large as tRNA and the prerequisites for the synthesis of proteins in a primitive form, utilizing a limited number of (sufficiently commonly occurring) amino acids?

XII. The Logic of Primordial Coding

XII. 1. The RRY Code

A most appealing speculative model for the origin of template-directed protein synthesis, recently proposed [3], is based on a number of logical inferences that are related to the problem of comma-free and coherent read-off. A primordial code must have a certain frame structure, otherwise a message cannot be read off consistently. Occasional phase slips would produce a frame-shifted translation of parts of the message and thereby destroy its meaning. The authors therefore propose a particular base sequence to which all codons have to adhere. Or, in other words, only those sequences of nucleotides that resemble the particular pattern could become eligible for messenger function. Uniformity of pattern could arise through instruction conferred by the exposed anticodon loop of tRNAs as well as by internal self-copying. Among the possible patterns that guarantee nonoverlapping read-off, the authors chose the base sequence purine-purine-pyrimidine, or, in the usual notation, RRY, to be common to all codons specifying a message. The particular choice was biased by a sequence regularity found in the anticodon loop of present tRNAs, which reads 3′NR$\alpha\beta\gamma$UY, $\alpha\beta\gamma$ being the anticodon, N any of the four nucleotides, and R and Y a purine and a pyrimidine, respectively. Another prerequisite of ribosome-free translation is the stability of the complex formed by the messenger and the growing polypeptide chain. A peptidyl-t-RNA must not fall off before the transfer to the subsequent aminoacyl-tRNA has been accomplished, that is, until the complete message is translated. Otherwise, only functionally inefficient protein fragments would be obtained. It is obvious from known base-pair stabilities that a simple codon-anticodon interaction does not guarantee the required stability of the messenger-tRNA complex. Therefore the model was based essentially on three auxiliary assumptions.

1. The structure of the anticodon loop of the adaptor (tRNA precursor) is such, that – given the particular and common codon pattern – an RNA can always form five base pairs with the messenger. The primitive tRNA is then assigned the general anticodon-loop sequence

where YYR is the anticodon.

2. The anticodon loop of each primitive tRNA can assume two different conformations, which are detailed in Figure 48. Both configurations had been described in an earlier paper by C. Woese [60] who

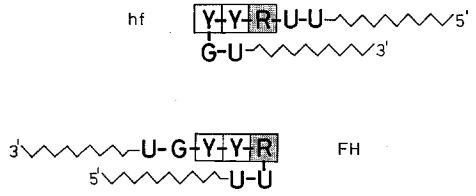


Fig. 48. Two possible configurations of the anticodon loop of tRNAs: FH according to Fuller and Hodgson [61] and hf according to Woese [60]. The anticodon pattern (framed) refers to the model of Crick et al. [3]

named them FH and hf. (FH refers to Fuller and Hodgson [61] who originally proposed that five bases at the 3′-end of the unpaired seven-base sequence in the anticodon loop are stacked on top of each other, while hf, according to Woese, designates a complementary configuration keeping the five bases at the 5′-end of the loop in a stacked position.) Woese assumed that the transition between both configurations plays an important role in ribosomal protein synthesis, but also referred to its possible significance in past, more primitive mechanisms.

3. Which of the two configurations is actually present depends on whether the tRNA is attached to an amino acid or to a peptide chain. By transferring an amino acid to the peptide chain, the tRNA flips from the hf to the FH configuration (cf. Fig. 49).

An additional, fourth postulate, not absolutely necessary as a prerequisite of the model, invokes an interaction between two adjacent tRNAs at the messenger, which assures that the required configration will be energetically the most favorable and contributes further to the stability of the polypeptide-messenger complex.

Figure 49 shows in more detail how polypeptide synthesis may be facilitated on the basis of the arguments given. The growing polypeptide chain is transported along the messenger, utilizing as 'fuel' the free energy of the transfer reaction. This reaction may be aided by a general nonspecific catalyst, but the mechanism does not require any sophisticated machinery, such as is nowadays provided by the ribosomes. Although interaction between codon and anticodon is stabilized by five base pairs, it is essential that the actual code make use of base triplets. It had been emphasized previously [62] that a primitive code utilizing anything but three bases is useless in explaining the present code.

The model explains well how in the absence of a sophisticated translation apparatus a message can be read off

spontaneously,
sequentially,
completely, i.e., unfragmented, and

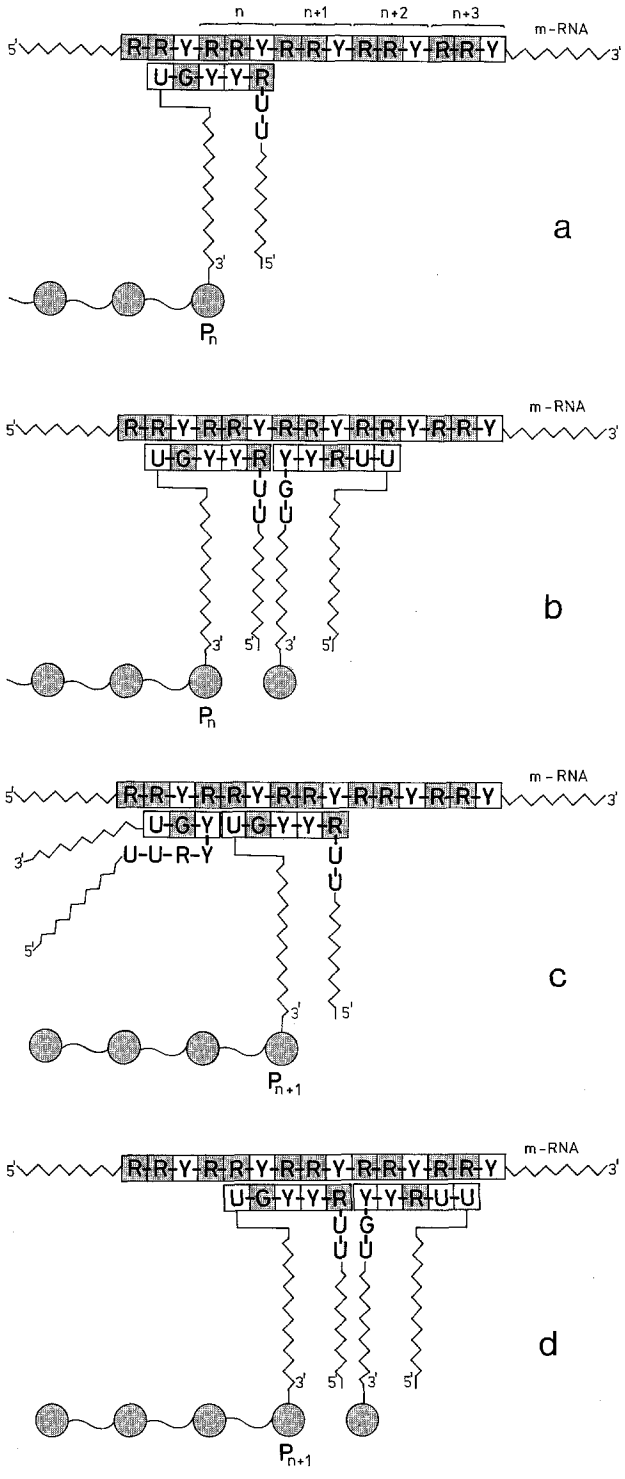


Fig. 49. The primitive translation mechanism requires 'sticky' interactions between the messenger and the peptidyl-tRNA. It thereby allows the growing peptide chain to remain in contact with the message until translation is completed. According to Crick et al. [3] the transport is effected by a flip mechanism involving conformational changes of the tRNA (FH⇌hf). The nascent peptide chain is always connected with the messengers via five base pairs with some additional stabilization by the adjacent aminoacyl-tRNA. The partial overlap of base pairing guarantees a consistent reading of a message encoded in base triplets

Naturwissenschaften 65, 341 – 369 (1978)

reproducibly, i.e., strictly maintaining a given codon frame.

The code is inherently related to certain structural features of the anticodon loop of present tRNAs, suggesting that these molecules are the descendents of the first functionally organized entities. The four amino acids assigned by this model are:

GG^{U}_{C} ;	GA^{U}_{C} ;	AG^{U}_{C} ;	AA^{U}_{C} ;
glycine	aspartic acid	serine	asparagine

Some of which were very abundant in the primitive soup [63].

On the other hand, the model also introduces some difficulties. A uniform RRY sequence has a large excess of purine over pyrimidine and therefore does not easily lend itself to stable internal folding.

As a consequence, such sequences

are quite labile toward hydrolysis
(if present as single strands),
have a greater tendency to form duplices
(which do not easily replicate by primitive mechanisms),
lack internal symmetry, and
produce minus strands of a different general code pattern (i.e., 5′RYY).

XII.2. The RNY Code

Before going into a more detailed discussion of the points listed above, let us now offer an alternative model that is free of these particular difficulties. The suggestion of using the general codon pattern RNY, where N stands for any of the four nucleotides A, U, G, C, is also to be credited to Crick et al. [1]. However, it was disfavored by its authors on the grounds of a disadvantage: If N represents a pyrimidine, then the anticodon loop, having the general sequence 3′ⓊGYNRUⓊ, can in some cases use only its five central nucleotides to form stable base pairs with the messenger. This argument may, however, be counteracted by the observation that an RNY code can assign eight amino acids, so that one may exclude certain combinations that do not fulfil the stability requirements for the messenger-peptidyl-tRNA complex (cf. below).

What are the advantages of a general code pattern RNY? First of all, the RNY code, like its RRY analog, is comma-free. Moreover, it is symmetric with respect to plus and minus strands. If read from 5′ to 3′, the general frame structure of both the plus and the minus strand is $R^{N'}_{N}Y$ where N and N′ are complementary, situated at mirror-image positions in the sequences of both strands. Similar symmetries can also develop internally within a single strand, allowing the formation of symmetric secondary folding structures. Typical examples of (almost) symmetric foldings are the present-day tRNAs. Single-stranded phage genomes and their derivatives (such as the midi-variant of Qβ-RNA) are also distinguished by such elements of symmetry. Here the selective advantage of a symmetric structure is obvious. If the molecules are to be reproduced by a polymerase, which recognizes specifically some structural feature, only a symmetric structure would allow the plus and the minus strands to be equally efficient templates. Such an equivalence of efficiency is required for selection. Thus the symmetry of tRNA may well be a relic from a time when these molecules still had to reproduce autonomously.

Internal folding also enhances the molcule's resistance to hydrolysis and offers an easy way of instructing the correct read-off of the message. In an open structure in which many nucleotides remain unpaired (e.g., for RRY sequences), replication and translation could start at any unpaired position of the sequence, leading to fragmentary products. In a completely paired structure, unmatched sticky ends are predestined to be the starting-points of replication and translation. In this way the complete message can be read off, requiring only transient partial unfoldings of the template structure, which may be enforced by interactions with the growing chain. Symmetry, although not absolutely required, would in this case again be of advantage (cf. Fig. 50).

From a logical point of view the RNY code seems to be more attractive than the RRY code, on the basis of three arguments.

1. Selective enhancement of RNA molecules must be effective for the plus as well as the minus strand. Symmetric RNY patterns can fulfil this requirement more easily than RRY sequences, which differ from their minus strands (RYY) and hence cannot both become equivalent targets for specific recognition by enzymes.

2. In view of the high complexity, there is little chance finding the very few sequences that offer useful properties for replication and translation. If these sequences, being symmetric structures, fulfil the requirements listed under (1), both the plus and the minus strands may be candidates for representing such functions.

3. Evolution of the translation apparatus with its various tRNAs and messengers requires a mutual stabilization of all replicative molecules. As will be shown below, hypercycles may emerge more easily from a quasi-species if this, on account of its symmetry, offers two complementary functions.

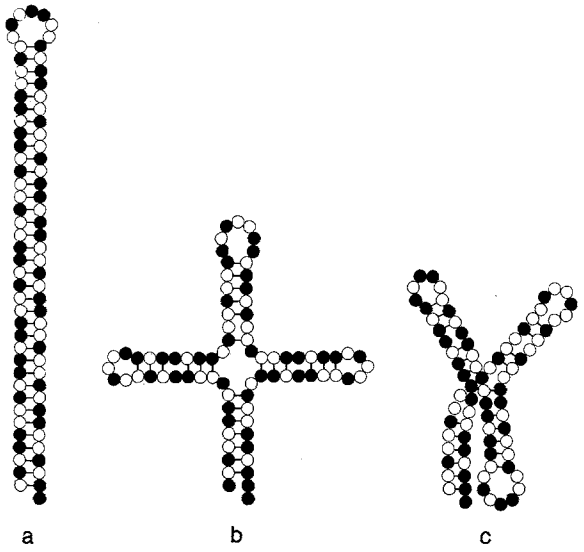


Fig. 50. Symmetric secondary structures of RNA require a corresponding internal complementarity (cf. Fig. 14 in Part A). Plus and minus strands will then exhibit similar foldings. RNY-codon patterns are able to produce such structures. A game has been devised, the rules of which take into account the physical interactions observed with oligonucleotides and tRNAs [17, 18]. It shows which of the structures is most likely to occur. Hairpins require two complementary halves of the molecule, e.g., a 5′-RRY sequence linked up to its minus strand involving a 5′RYY pattern

A system can accumulate information and eventually evolve to higher complexity only if it adopts the 'Darwinian logic' of selective self-organization. However, this logic must find its basis and its expression in material properties. All that the constituents can recognize at the beginning is natural abundance and strength of interaction. These are the properties with which we shall have to deal in order to understand the start of translation.

XIII. Physics of Primordial Coding

XIII.1. The Starting Conditions

Self-organization as a multimolecular process requires monomers as well as polymers to be present in sufficiently high concentrations. Its onset therefore must have been preceded by an extended phase of prebiotic synthesis, during which all the material necessary for yielding a 'highly enriched soup' was accumulated. We do not intend to dwell on these processes of primordial chemistry, nor do we quarrel about details of historical boundary conditions. Questions as to whether the 'soup' originated in the oceans, in a pond, or even in small puddles, or whether interfaces, coarse-grained, or porous surfaces were involved, may be important if absolute rates of the *historical* processes are to be estimated.

Here we simply start from the assumption that when self-organization began all kinds of energy-rich material were ubiquitous, including in particular:
amino acids in varying degrees of abundance,
nucleotides involving the four bases A, U, G, C,
polymers of both preceding classes, i.e., proteinoids as well as tRNA-like substances, having more or less random sequences.

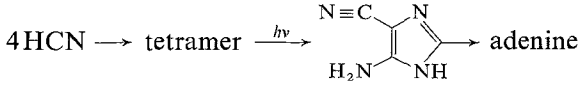
'Less random' in this context means the existence of nearest-neighbor and more complex folding interactions, leading to a preference of certain structures, while 'more random' refers to their primary unrelatedness to any functional destination, which – if initially present – could only be coincidental.

On the other hand, we definitely do not suppose the presence of any adapted protein machinery, such as specific polymerases,
adapted synthetases, or any of the ribosomal functions.

This does not exclude an involvement of non-instructed, poorly adapted protein catalysts, in facilitating the start of replication and translation. However, not being able to reproduce and improve selectively, those proteins must be subsumed together with other catalytic surfaces under 'constant environmental factors'.

XIII.2. Abundance of Nucleotides

Since nucleic-acid-like structures are the only ones that offer inherent self-reproductive properties, it is important to analyze first their abundances as well as their mutual interactions in more detail.
The monomeric nucleotides, especially their energy-rich oligophosphate forms, are more difficult to obtain (using possible prebiotic mechanisms) than are amino acids. Quantitative statements of relative abundance are therefore scarce. S. Miller and L. Orgel ([63], p. 104) emphasize the central role of adenine nucleotides both in genetic processes as well as in energy transfer and correlate it with the relative ease with which this substance is formed. J. Oró and his co-workers [64] found that adenine can be obtained in yields of 0.5% in concentrated aqueous solutions of ammonium cyanide, while Miller and Orgel ([63], p. 105) showed that even hydrogen cyanide alone, in a reaction catalyzed by sunlight, yields the important intermediate



Naturwissenschaften 65, 341 – 369 (1978)

The same intermediate can also react with cyanate, urea, or cyanogen to give guanine. Less well understood are the mechanisms of pyrimidine synthesis. A pathway could be demonstrated for the synthesis of cytosine, using cyanoacetylene – an electric-discharge product of mixtures of methane and nitrogen – in combination with cyanate. Uracil, on the other hand, appears to be a hydrolysis product of cytosine and may possible owe its primordial existence to this source.

Only little can be said about the primordial natural abundance of the purines and pyrimidines. The rate of template-instructed polymerization is proportional to the concentration of the monomer to be included. For complementary instruction, at least two nucleotides are required and they will have to be equivalently represented in informational sequences. Therefore the inclusion of the less abundant nucleotide will always be the rate-limiting step, at least for chain elongation. A possible large excess of A over U in the primordial monomer distribution would then have been of little help in favoring AU over GC copolymers, except in cases where the nucleation of oligo-A primers is the rate-limiting step. The capacity for replicative growth is limited to the template function of the less abundant member of the complementary nucleotide pair. If under primordial conditions the abundance of G and C was intermediate to that of A and U, then GC- and AU-rich copolymers may well have formed with comparable rates.

We therefore cannot maintain the earlier, speculative view that the first codons were recruited exclusively from a binary alphabet, made up of AU copolymers alone.

XIII.3. Stability of Complementary Structures

The stability of base pairing may provide clues that are more illuminating with respect to the question of the first codons. Stabilities and rates of base pairing have been studied using various nucleotide combinations. These results have been discussed in detail in earlier reviews [44, 4]. They reveal quantitatively the generally accepted view that GC pairs within a cooperative stack provide considerably more stability than AU pairs.

The stability constant of a continuous and homogeneous oligomeric sequence of n nucleotide pairs can be represented by the relation

$$K_n = \beta s^n \tag{91}$$

which refers to a linear Ising model. The factor β is a cooperativity parameter, which for both the AU and GC pairs amounts to an order of magnitude of 10^{-3}*, while s is the stability constant of a single pair within the cooperative stack. For homogeneous sequences of AU pairs this parameter s is about one order of magnitude smaller than for homogeneous sequences of GC pairs, or in a rough quantitative representation:

$$s_{AU} \approx 10 \quad \text{while} \quad s_{GC} \approx 100.$$

Higher absolute stabilities than predicted by relation (91) are found if one of the complementary strands can assume the particular stacking configuration present in the anticodon loop of tRNA. Presumably the cooperativity parameter β is changed in this case. Yet O.C. Uhlenbeck, J. Batter and P. Doty [65, 66] found that tri- and tetranucleotides, complementary to the anticodon region of a tRNA and differing in one AU pair, exhibit a difference of one order of magnitude in their stability constants, quite in agreement with the figure quoted above. It is also reasonable that the largest absolute values of stability constants found are those for the interaction of two tRNAs that are complementary in their anticodons [67].

The data obtained with defined short sequences may serve at least for a comparison between various replication and translation models and for conclusions about their relative significance. It is obvious that single isolated AU or GC pairs are unstable under any realistic conditions of concentration. The start of replication, therefore, requires special help in the form of primer formation, and it is particularly this step that calls first for enzymic support. Present-day phage RNA replicases are also specifically adapted to a primary sequence pattern of the phage genome. For chain elongation, the incoming nucleotide is bound cooperatively on top of a stack of base pairs of the growing chain. Here the data suggest that the GC pair is about ten times more stable than the AU pair, resulting in a relatively higher fidelity q for G and C than for A and U. If the rate of replication is limited by the formation of the covalent link in the polynucleotide backbone (rather than by base-pair formation) the fidelity can be correlated with the monomer concentrations m_R and m_Y and the pair-stability constants K_{RY}, K_{RR}, and K_{YY}. The reproduction fidelity for any given nucleotide then may be obtained from the geometric mean of the fidelities for both complementary processes,

$$\text{R}\rightarrow\text{Y}; \quad \text{Y}\rightarrow\text{R}:$$

$$q_{RY} = \frac{m_Y K_{RY}}{\sum_N m_N K_{RN}} \quad \text{and} \quad q_{YR} = \frac{m_R K_{YR}}{\sum_N m_N K_{YN}} \tag{92}$$

* Such a relation is formally valid for both the internal base pairing within a given sequence and the binary association of two complementary sequences, where β has the dimension M^{-1}.

where $K_{RY} \approx K_{YR}$ and summation is extended over all N = A, U, G, C. Those q values are identical for A and U or G and C, respectively, since the error can appear either in the plus or in the minus strand. If the monomeric concentrations are of equal magnitudes, the stability constants determine what fidelities are obtainable. Then it follows that G and C reproduce considerably more accurately than A and U. The ratio of the error rates for GC and AU reproduction in mixed systems, however, does not exactly resemble the (inverse) ratio of the corresponding stability constants, owing mainly to the presence of GU wobble interactions, which are the main source of reproduction errors – even in present-day RNA phage replication [34].

We have made a guess of q values based on various sets of data for enzyme-free nucleotide interactions. They are summarized in Table 15. The first three sets refer to equal monomeric concentrations of A, U, G, and C. This assumption may be unrealistic and is therefore modified in the last three examples. One may object to the application of stability data that were obtained from studies with oligonucleotides. However, the inclusion of a single nucleotide in the replication process involves cooperative base-pair interactions and hence should resemble the relative orders found with oligonucleotides. All that is required for calculating the q values are *relative* rather than *absolute* stabilities.

The conclusion from the different estimates presented in Table 15 is: G and C reproduce with an appreciably higher fidelity than A and U. Depending on the superiority of the selected sequences (σ, cf. Eq. (28), Part A), the reproducible information content of GC-rich sequences in early replicative processes is limited to about twenty to one hundred nucleotides, i.e., to tRNA-like molecules, while that of AU-rich sequences can hardly exceed ten to twenty nucleotide residues per replicative unit. It should be emphasized at this point, that longer sequences of *any* composition may well have been present. However, they were *not* reproducible and therefore could not evolve according to any functional requirement.

From an analysis of experimental data for phage replication we concluded in Part A that even well-adapted RNA replicases would not allow the *reproducible* accumulation of more than 1000 to 10000 nucleotides per strand. This is equivalent to the actual gene content of the RNA phages.

We may now complete our statement regarding primordial replication mechanisms: A size as large as tRNA is *reproducibly* accessible only for GC-rich structures. Hence:

GC-rich sequences qualify as candidates for early tRNA adapters and for reproducible messengers, at least as long as replication is not aided by moderately adapted enzymes.

A similar conclusion can be drawn with respect to the start of *translation.* As was emphasized by Crick et al. [3], stability of the peptidyl-tRNA-messenger complex is critical for any primitive translation mo-

Table 15. Estimates of fidelities and error rates for G and C vs. A and U reproduction

Monomer concentrations	Stability constants of base pairs	Fidelity q		Error rate $1-q$	
		GC	AU	GC	AU
$m_A = m_G = m_C = m_U$	$K_{RR} = K_{YY} = 1$ $K_{AC} = 1$; $K_{GU} = 10$ $K_{AU} = 10$; $K_{GC} = 100$	0.93	0.59	0.07	0.41
$m_A = m_G = m_C = m_U$	$K_{RR} \approx K_{YY} \ll 1$ $K_{AC} = 1$; $K_{GU} = 10$ $K_{AU} = 10$; $K_{GC} = 100$	0.95	0.67	0.05	0.33
$m_A = m_G = m_C = m_U$	$K_{RR} = K_{YY} \ll 1$ $K_{AC} = 1$; $K_{GU} = 5$ $K_{AU} = 10$; $K_{GC} = 100$	0.97	0.78	0.03	0.22
$m_A = 10 m_G$ $m_G = m_C$ $m_C = 10 m_U$	$K_{RR} \approx K_{YY} \ll 1$ $K_{AC} = 1$; $K_{GU} = 5$ $K_{AU} = 10$; $K_{GC} = 100$	0.93	0.81	0.07	0.19
$m_A = 10 m_G$ $m_G = m_C$ $m_C = 10 m_U$	$K_{RR} \approx K_{YY} \ll 1$ $K_{AC} = 1$; $K_{GU} = 5$ $K_{AU} = 10$; $K_{GC} = 100$	0.95	0.69	0.05	0.31
$m_A = 10 m_G$ $m_G = m_C$ $m_C = 10 m_U$	$K_{RR} = K_{YY} = 1$ $K_{AC} = 2$; $K_{GU} = 10$ $K_{AU} = 10$; $K_{GC} = 100$	0.86	0.25	0.14	0.75

Naturwissenschaften 65, 341 – 369 (1978)

del. Applying the data quoted above, the stability constant of a complex including five GC pairs amounts to

$K_{5GC} \approx 10^7\ M^{-1}$

while that for five AU pairs is five orders of magnitude lower:

$K_{5AU} \approx 10^2\ M^{-1}$

Again these values must be seen as relative; they might actually be somewhat larger if the stacked-loop region or tRNA (as we know it today) were involved, which, however, would not invalidate the argument based on relative magnitudes.

We might also evaluate the models on the basis of lifetime data. The recombination-rate constants, as measured for complementary chains of oligonucleotides, were found consistently to lie in the order of magnitude of

$k_R \approx 10^6\ M^{-1}\ s^{-1}$.

Given the stability constants quoted above, the lifetimes of the respective complexes would amount to

$\tau_{5GC} \approx 10\ s$ and $\tau_{5AU} \approx 10^{-4}\ s$.

Again, these numbers might shift to larger values if stabilities turned out to be higher, and if two adjacent tRNAs are able to stabilize each other when attached to the messenger chain. Then lifetimes might just suffice for GC-rich sequences to start primitive translation. The lifetimes are certainly much too short if AU pairs are in excess. We see now that the slight disadvantage of the RNY relative to the RRY code resulting from stabilities can be balanced by utilizing primarily G and C at least for part of the R and Y positions. A four-membered GC structure is definitely more stable than any five-membered structure, including more than two AU pairs.

The conclusion is:

The start of translation is highly favored by GC-rich structures both of the tRNA precursors and of the messengers.

XIV. The GC-Frame Code

XIV.1. The First Two Codons

If we combine the conclusions drawn from stability data with the arguments produced by Crick et al., we can predict which codon assignments were probably the first ones.

The only sufficiently long sequences that are able to reproduce themselves faithfully must have been those in which G and C residues predominated. The first codons were then exclusively combinations of these two residues. The requirement for a comma-free read-off excludes the symmetric combinations GGG/CCC and GCG/CGC. This may be easily verified by writing down such sequences. Adaptors with the correct anticodon combinations can bind in various overlapping positions. This will have even more deleterious consequences, if further codon combinations, derived from symmetric precursors, are introduced. We are thus left with two complementary pairs of combinations, namely, GGC/GCC and CCG/CGG (all patterns to be read from 5′ to 3′). From the point of view of symmetry both appear completely equivalent. There is, however, a slight asymmetry based on wobble in the third position. Let us compare messenger sequences consisting exclusively of either CNG or GNC codons. In the first case the wobble base G is always situated in the third codon position, while in the second case it is restricted to the first position, in both the plus and the minus strands (if consistently read from 5′ to 3′). For replication the different codon positions are not distinguishable. Hence, wherever a wobble base occurs, an ambiguity may be introduced in the complementary strand, which, when it comes to translation, in one case affects the first, in the other case the third codon position:

5′----CNG CNG CNG----3′

↓

5′----$\overset{C}{U}$N′$G\overset{C}{U}$N′$G\overset{C}{U}$N′G----3′

and

5′----GNC GNC GNC----3′

↓

5′----$GN\overset{C}{U}GN\overset{C}{U}GN\overset{C}{U}$ ----3′

Only in the second case are the reproduced sequences correctly translated, i.e., if wobble interactions with the adapter occur preferentially in the third rather than in the first codon position.

In other words, an adapter with the anticodon 3′CNG can read both 5′GN′C and 5′GN′U, whereas an adapter with the anticodon 3′GNC will only read 5′CN′G, but reject 5′UN′G. The argument might be weak if five base pairs are required to keep the adapter bound to the messenger, since then wobble positions may not be as clearly distinct. Nevertheless, this asymmetric relationship between the first and third codon positions does exist and is obvious in the present genetic code.* Relatively small selective advantages are usually sufficient to bias the course of evolution. Crick et al.

* Our argument is aided by the fact that in the stationary distribution G is more persistent than C.

obviously preferred the RRY (or RNY) model on the basis of such arguments, too.
We are now able to make a unique assignment for the first two codons, namely

5′GGC and 5′GCC

which are complementary if aligned in an antiparallel fashion. This choice was dictated by four arguments, viz.,

stability of adapter-messenger interaction and
fidelity of replication, both favoring GC combinations to start with,
comma-free read-off in translation requiring an unsymmetric GC pattern, and
consistency of translation restricting *wobble* ambiguities to the third codon position.

We should like to emphasize that these arguments are based exclusively on the properties of nucleic acids. It is satisfying to notice that the two codons GGC and GCC in the present genetic code refer to the two simplest amino acids, glycine and alanine, which in experiments simulating primordial conditions indeed appear with by far the greatest abundance.
One may object that translation products consisting of these two residues only, will hardly represent catalysts of any sophistication. We shall return to this question in Section XVI. At the moment it suffices to note that translation at this stage is not yet a property required for the conservation of the underlying messengers. The first GC-rich strands are selected solely on the basis of structural stability and their ability to replicate faithfully. Many different GC sequences would serve this purpose equally well and hence may have become jointly selected as (more or less degenerate) partners of one quasi-species. Symmetric structures are greatly favored here, because they can fulfil the criteria of stability for the plus and for the minus strand simultaneously.
Among stable members, perhaps induced by template function of anticodon loops and subsequent pattern duplications, comma-free code sequences may have occurred and then started translation. If their translation products add any advantage to the stability or to the reproduction rates of their messengers, they will evolve further by a Darwinian mechanism and thereby change continuously the quasi-species distribution. Before we come back to such a stabilization by means of translation products, let us enlarge somewhat more on the question of stability of structure versus efficiency of replication, because it seems that both required properties are based on conflicting prerequisites.

XIV.2. The 'Aperiodic Linear GC Lattice'

The tRNA-like molecule with its internal folded structure strengthened by hydrogen bonds may be considered a microcrystallite. If it involves longer stretches of complementary GC pattern, the resulting internal structure may be quite inert. From melting curves of tRNA loops or corresponding oligonucleotides we know that an uninterrupted sequence of only four GC pairs is already quite stable. S. Coutts [68] studied an oligonucleotide corresponding to the extra loop of $tRNA^{Ser}_{1,2}$ from yeast, which contains 4 GC pairs (and was obtained by partial digestion of the tRNA molecule). He found a melting temperature of $84 \pm 1°$ C and a ΔH of 44 ± 4 kcal/mol. This is equivalent to a stability constant of about 2×10^5 at 25° C, in good agreement with the figures mentioned above. What is rated in the *selection* of a quasi-species is not merely the structural stability of the strands, but rather a favorable combination of structural stability with reproductive efficiency. Efficient template function requires a quick partial unwinding of a loop, a procedure for which excessively long stretches of GC pairs are prohibitive. However:

Natural sequences are not perfect anyway.

Given a high abundance of A-monomers and the limited fidelity of base pairing, the GC microcrystallites will always be highly doped with A-residues, acting like imperfections in the linear GC lattice. A priori, there may be any kind of sequence from high to low A, U, G, or C content. What is to be selected and then reproducibly mulitplied, will be a sequence rich in GC, but not perfect. If, for instance, every fifth position in such a sequence is substituted by an A or U residue, then base-paired regions, depending on internal complementarity, will involve on the average no more than four GC pairs (cf. present tRNA). Those structures can melt locallly with ease, especially if the replication process is aided by a protein, which then represents the most primitive form of a replicase.

Note: *A-U imperfections in the aperiodic GC lattice are selectively advantageous.*

As Thomas Mann* said: 'Life shrinks back from absolute perfection.'

XIV.3. From GNC to RNY

Given a certain abundance of A (and complementary U) imperfections in the GC-rich strands of the

* Th. Mann: Der Zauberberg (The Magic Mountain)

Naturwissenschaften 65, 341 – 369 (1978)

selected quasi-species, the next step in the evolution of a code seems to be preprogrammed. Mutations might occur in any of the three codon positions, but their consequences are quite different. Substitution of the middle base of a codon would enforce a complementary substitution of the middle base of the corresponding anticodon occurring in the minus strand and hence immediately introduce two new codons, GAC and GUC. Changes in the first or third position, on the other hand, would be complemented by changes in the third (or first) position, respectively, of the minus strand and – by wobble arguments – finally lead to only one further assignment. Moreover, the GC frame for comma-free reading would be perturbed.

Stability requirements do not initially allow for a substitution of more than one AU pair in the five-base-pair region of the messenger-tRNA complex. Hence the most likely codons to occur next are 5′GAC and 5′GUC. Being mutants of the pre-existing pair 5′GGC/5′GCC, they may be abundantly present as members of the selected GC-rich quasi-species.

However, if these mutants are assigned a function in translation, they have to become truly equivalent to the dominant 5′GGC/5′GCC species. It is at this stage that hypercyclic stabilization of the four codon adapters (and the messengers which encode the coupling factors) becomes an absolute requirement. Without such a link the different partners of the primary translation system may coexist for some time, but they would never be able to evolve or to optimize their cooperation in any coherent fashion.

The four codons allow four different amino-acid assignments, which can now offer a rich palette of functions. The resulting proteins therefore could become efficient coupling factors. Messengers and tRNAs, as members of the same quasi-species, might have emerged from complementary strands of the same RNA species, thus sharing both functions.

On the other hand, this may be too restrictive a constraint for their further evolution. We then have to assume that they were derived from a common precursor, but later on diverged into different seqences owing to their quite different structural and functional requirements.

The assignments for GAC and GUC, according to the present table of the genetic code, are aspartic acid and valine. Before we discuss the amino-acid aspect in more detail we may look briefly at some further steps in the evolution toward a more general code.

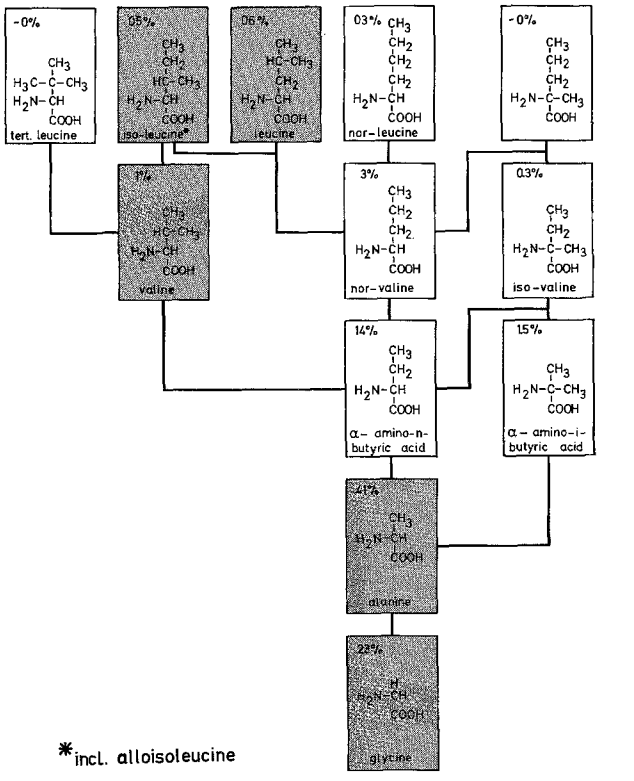
High stability of codon-anticodon interaction is required less and less as the translation products become better adapted. Wobble interactions are finally admitted and the GC frame code can evolve to the more general RY frame code. All together this brings four more amino acids onto the scene. The first substitution still occurs under the stability constraint, which forces the AU content to be as low as possible. Hence it introduces the two codons 5′AGC (=serine) and 5′ACC (=threonine). Their complementary sequences affect the third codon position, yielding 5′GCU and 5′GGU, which reproduce the assignments for alanine (GCC) and glycine (GGC). The degeneracy, accounting for the wobble interactions in the reproduction of these latter codons, may have been the primary cause of the appearance of AGC and ACC codons and their assignments.

If with the evolution of an enzymic machinery more than one AU pair is allowed in the codon region we arrive at two more new assignments, namely, AA^{U}_{C}(asparagine) and AU^{U}_{C} (isoleucine). This completes all possible assignments for an RNY code. The further evolution of the genetic code requires a relaxation of the constraint of nonoverlapping frames. Adaptation of ribosomal precursors is therefore now imperative.

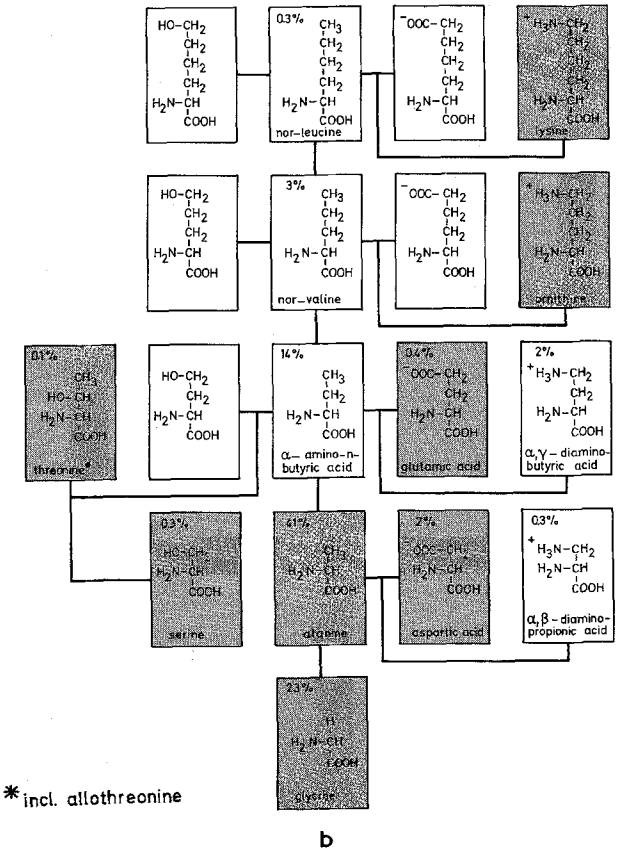
XIV.4. The Primary Alphabet of Amino Acids

Quite reliable estimates can be made for the *primordial abundance* of various amino acids. Structure and composition already provide the main clues for a guess about the likelihood of synthesis under primordial conditions. In Figure 51 the family tree of the first dozen nonpolar aliphatic amino acids as well as a few branches demonstrating the kinship relations for the simpler polar side chains are shown. Interesting questions concerning Nature's choice of the protein alphabet arise from this diagram.

The two simplest amino acids, glycine and alanine, are 'natural' representatives. It was apparently easier to fulfil requirements for hydrophobic interaction by adding some of the higher homolog, such as valine, leucine, and iso leucine. This specific choice may have been subject to chance; perhaps it was biased also by discriminative interactions with the adapters available. Among the polar side chains we find some obvious aliphatic carboxylic acids (aspartic and glutamic acid) as well as alcohols (serine and threonine), but not the corresponding amines (α, β-diaminopropionic acid and α,γ-diaminobutyric acid). Only the second next homolog (lysine) appears among the twenty 'natural' amino acids, while the intermediate (ornithine) still shows its traces. The reason may be that upon activation of the second amino group lactame formation or elimination occurs, which terminates the polymerization. Moreover, the second amino group may lead to a branching of the polypeptide chain (although



a



b

a similar argument may be raised for the carboxylic groups). Positively charged side chains may well have been dispensible in the first functional polypeptides. Even in present sea water the concentration of Mg^{2+} is high enough (~ 50 mM) to cause an appreciable complexation with carboxylic groups. Under reducing conditions even more divalent ions (such as Fe^{2+}) may have been dissolved in the oceans. Those metal ions, attached to carboxylic groups and still having free coordination sites, are especially important for close interactions between early proteins and (negatively charged) polynucleotides. From this point of view, side chains containing negatively charged ligands seem to be less dispensible than those containing positive charges.

The 'natural' amino acids not appearing in Figure 51 bear considerably more complex sidechains and were therefore present in the primordial soup at comparatively low concentrations.

These guesses from structure and composition are excellently confirmed by experiments simulating the prebiotic synthesis of amino acids, carried out by S. Miller and others (reviewed in [63]). The yields obtained for the natural amino acids (but also for other branches of the family tree) correspond grossly to the chemist's expectations (cf. numbers in Fig. 51), although many interesting items of detailed information are added by these experiments. The results are, furthermore, in good agreement with data obtained from meteorite analysis [69, 70] reflecting the occurrence of amino acids in interstellar space. Table 16 contains a compilation of data (taken from [63]), which are relevant for our discussion.

There is no doubt that the primordial soup was very rich in glycine and alanine. In Miller's experiments these amino acids appear to be about twenty times more frequent than any of the other 'natural' representatives. The next two positions in the abundance scale of natural amino acids are held by aspartic acid and valine with a clear gap between these and leucine, glutamic acid, serine, isoleucine, threonine, and proline.

There is every reason to assume that assignments of codons to amino acids actually followed the abundance scale. If glycine and alanine are by far the most abundant amino acids, why should they not have been assigned first, as soon as chemical mecha-

Fig. 51. The family tree of the first aliphatic amino acids and some branches for the simplest polar side chains. The number in the left upper corner of each plate refers to Miller's data of relative yields under primordial conditions [63] (i.e., molar yield of the particular amino acid divided by the sum of yields of all amino acids listed in their Table 7–2 on p. 87). The plates for the natural amino acids are shaded

Table 16. Abundance of natural amino acids in simulated prebiotic synthesis and in the Murchison meteorite. The first column refers to those amino acids which appear in the proteins. The data in the second column are typical results from S. Miller's experiments (as reviewed in [63]). They were obtained by sparking 336 mMol of methane in the presence of nitrogen and water. The total yield of amino acids (including those which do not appear in the proteins) based on carbon was 1.9%, the corresponding yields of glycine and alanine 0.26% and 0.71%, resp. The yields in the table refer to molar abundance. Similar data are obtained under different conditions, gly, ala, asp, val usually occurring as the most abundant natural amino acids. The meterorite data were reported by J. Oró et al. [69] and by K.A. Kvenvolden et al. [70], D-isomers appeared in all cases to be close to 50%. Further literature can be found in [63]

Compound	Yield [μM]	μg/g meteorite
Glycine	440	6
Alanine	790	3
Aspartic acid	34	2
Valine	19.5	2
Leucine	11.3	
Glutamic acid	7.7	3
Serine	5.0	
Isoleucine	4.8	
Threonine*	1.6	
Proline	1.5	1

* Including allothreonine

nisms of activation became available? The first primordial polypeptides – the instructed as well as the noninstructed – must then have been largely gly-ala copolymers with only occasional substitutions by other amino acids, probably including also those which finally did not become assigned.

The agreement between the abundances of natural amino acids and the order of the first four codon assignments is striking. It should be emphasized that our choice of codons is based exclusively on arguments that are related to the structure of nucleic acids. Not only do the first four GC frame codons coincide exactly with the order of abundance, but furthermore the four additional RNY codons are assigned to amino acids which – with the exception of asparagine – are well represented with appreciable yields in Miller's table. One may well ask whether the assignment of AA_C^U to asparagine is a primary one or – as suggested by the intimate neighborhood to the lysine codon – was originally related to any of the lower diamino acid homologs, which appear in Miller's table with a reversed order of yields as compared with aspartic acid (α,γ-diaminobutyric acid) and glutamic acid (α,β-diaminopropionic acid). Without further evidence, however, this might be a too far-reaching speculation.

Also of importance in this respect are some recent results from the protein-sequence analysis of nucleotide-binding enzymes, which are believed to have existed more than $3 \cdot 10^9$ years ago under precellular conditions [71, 72]. The data suggest the existence of a precursor sequence of the nucleotide-binding surface, which include the amino acids valine, aspartic (and glutamic) acid, alanine and glycine besides isoleucine, lysine and threonine (although these data actually refer to a later stage of precellular evolution than is discussed in this paper).

XV. Hypercyclic Organization of the Early Translation Apparatus

Any model for the evolution of an early code and translation apparatus will have to provide conditions that allow tRNA-like adapters as well as gene precursors (or messengers) for various enzymic factors not only to coexist, but also to grow coherently and to evolve to optimal function. In Parts A and B it was shown that such a self-organization requires cyclically closed reactive links among all individual partners, unless they all can be integrated structurally into one replicative unit. In this section we have to show how realistic code models are correlated with an hypercyclic organization, and how such systems can evolve.

An apparent difficulty of the hypercycle is how it is to originate. In plain language, the abundant presence of all members of the hypercycle seems to be a prerequisite for its origin. In more scientific terms, the hypercycle, being a higher-order-reaction network, has to 'nucleate' by some higher-order mechanism in order to come into existence.

Consider, for comparison, a simple replicative unit that grows according to a first-order autocatalytic rate law. In a solution buffered with energy-rich building material one copy is sufficient to start the multiplication process. Those experiments have been carried out with phage RNA or its noninfectious variants [7, 8, 32, 34, 73]. One template strand is sufficient to produce – within a few minutes – a large population of identical copies (cf. Part A).

A hypercycle never could start in this way. A single template copy would not multiply unless a sufficiently large number of its specific catalytic correspondents were present. These in turn are encoded by templates which themselves could not have multiplied without the help of their translation products. The growth of all templates in the system is dependent upon catalytic support, but the catalysts cannot grow unless the templates multiply. How large is the probability that nucleation will occur through some accidental fluctuation? Let us assume a test tube with 1 ml of sample solution. Diffusion-controlled reactions of

macromolecules may have rate constants of the order of magnitude of 10^8 M^{-1} s^{-1}. Hence at least 10^8 identical copies of a given catalytic reaction partner have to be present in order to start template multiplication with a half-time of about one day. There is no chance that correlated functions among several such partners could result from coincidences of such giant fluctuations. It may, of course, be possible that the various templates multiply according to mixed-order terms; in other words: that first-order (enzyme-free) autocatalytic terms (representing template multiplication without catalytic help by other members) are superimposed upon the second-order catalytic replication terms. The hypercyclic link would then become effective only after concentrations have risen to a sufficiently high level. However, the system cannot know in advance which of the many alternative sequences multiplying according to a first-order autocatalysis are the ones which provide the useful information for the catalysts required at the later stages of organization.

There is only one solution to this problem:

The hypercycle must have a precursor, present in high natural abundance, from which it originates gradually by a mechanism of mutation and selection.

Such a precursor, indeed, can be the quasi-species consisting of a distribution of GC-rich sequences. All members of a stable quasi-species will grow until they are present in high concentrations. As was shown in Section XIV, some GC-rich sequences may be able to start a translation by assigning amino acids to defined anticodons. At this stage the translation products are really not yet necessary for conserving the system, so translation can still be considered a game of trial and error. If, however, it happens that one of the translation products offers advantages for the reproduction of its own messenger, this messenger may become the dominant representative of the quasi-species distribution.

A single RNA species could at best assign a two-amino-acid alphabet, if both the plus and the minus strands act as adapters for two complementary codons (e.g., GGC and GCC). If adapter sequences are sufficiently abundant, there is also a finite chance that coexisting mutants assign the two or even four codons (including GAC and GUC for aspartic acid and valine), again possibly utilizing both plus and minus strands. All this may still happen during the quasi-species phase.

Such a system, however, can evolve only if the different RNA species stabilize each other with the help of their translation products. We defer a discussion of the details of assignments – e.g., as to whether plus and minus strands of a given RNA species can evolve concomitantly and thereby become two adapters for complementary codons, or whether the plus strand as messenger encodes for the coupling factor, while only the minus strand acts as an adapter – to Section XVI. Here we study the problem of how hypercyclic organization can gradually evolve out of a quasi-species distribution.

Figure 52 shows how such a process can be envisaged. Assume two abundant mutants of the quasi-species, whose plus and minus strands are able to act as adapters of (at most) two amino acid pairs (e.g., gly/ala and asp/val), and which at the same time may be translated into a protein made up of (at most) four classes of amino acids. If the translation products offer any catalytic function in favor of the reproduction of their messengers, one would probably encounter one of the situations represented in Figures 52 or 53.

Both messengers, being closely related mutants, encode for two proteins with closely related functions. If one is a specific replicase, the other will be too, both functions being self- as well as mutually enhanc-

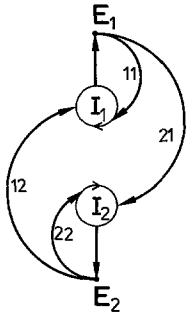


Fig. 52. *Two mutant genes I_1 and I_2*, encoding for their own replicases E_1 and E_2, may show equivalent couplings for self- [11, 22] and mutual [21, 12] enhancement due to their close kinship relation. Analogous behavior can be found in present RNA-phage replicases

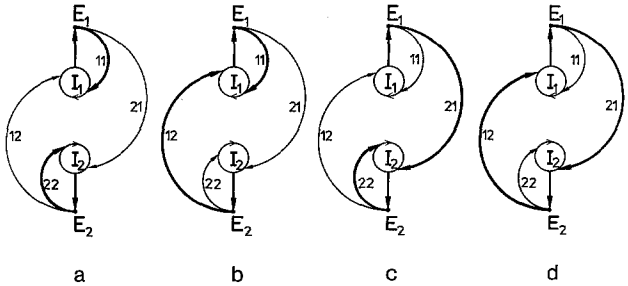


Fig. 53. *The evolution principle of hypercycles* is illustrated by the four possible situations arising from the couplings between two mutants shown in Figure 52. The thick lines indicate a preference in coupling (however small it may be). A stable two-membered hypercycle requires a preference for mutual enhancements as depicted in d)

Table 17. Fixed-point analysis of the two-member hypercycle

represented in Figure 52 has been carried out using the simplified rate equations

$$\dot{x}_i = \sum_{k=1,2} k_{ik} x_i x_k - \frac{x_i}{c} \sum_{l=1,2} \sum_{m=1,2} k_{lm} x_l x_m$$

$$i=1,2; \quad x_1+x_2=c$$

yielding the three fixed points and their eigenvalues:

$$\bar{\mathbf{x}}_1=(c,0); \quad \omega^{(1)}=(k_{21}-k_{11})c$$

$$\bar{\mathbf{x}}_2=(0,c); \quad \omega^{(2)}=(k_{12}-k_{22})c$$

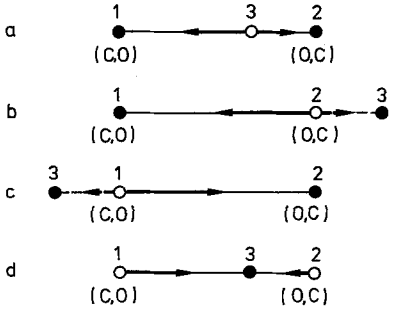
$$\bar{\mathbf{x}}_3=(k_{22}-k_{12}, k_{11}-k_{21})\frac{c}{k_{11}-k_{21}+k_{22}-k_{12}};$$

$$\omega^{(3)}=\frac{(k_{11}-k_{21})(k_{22}-k_{12})}{k_{11}-k_{21}+k_{22}-k_{12}}c.$$

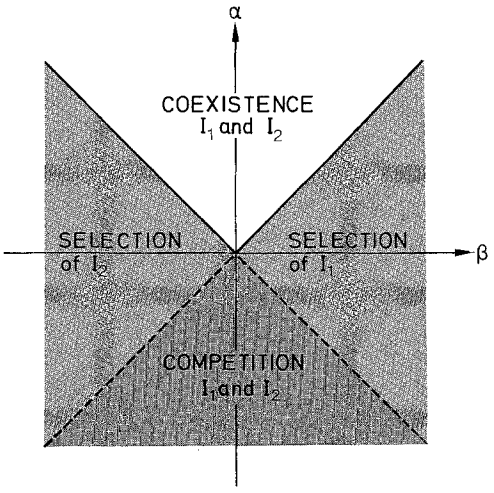
Four cases may be distinguished

a) $k_{11}>k_{21}$; $k_{22}>k_{12}$ yielding competition between I_1 and I_2
b) $k_{11}>k_{21}$; $k_{22}<k_{12}$ yielding selection of I_1
c) $k_{11}<k_{21}$; $k_{22}>k_{12}$ yielding selection of I_2
d) $k_{11}<k_{21}$; $k_{22}<k_{12}$ yielding hypercyclic stabilization of I_1 and I_2

The fixed-point diagrams of these four cases are (cf. Part B)



A unified representation can be achieved if the two coordinates: $\alpha=k_{12}+k_{21}-k_{11}-k_{22}$ and $\beta=k_{12}-k_{21}+k_{11}-k_{22}$ are introduced.

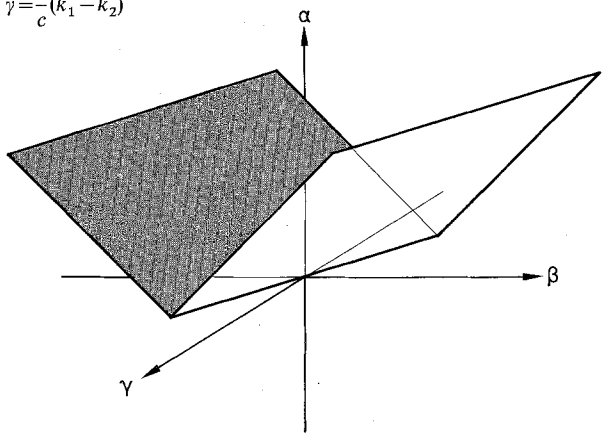


If in addition to the second-order term of the rate equations a linear autocatalytic term is introduced (yielding a growth function of the form $\Gamma_i=k_i x_i+\sum_{j=1,2} k_{ij} x_i x_j$), the region of stable hypercyclic coexistence of both species I_1 and I_2 is the space above the folded sheet in the three-dimensional parameter space with the coordinates:

$$\alpha=k_{12}+k_{21}-k_{11}-k_{22}$$

$$\beta=k_{12}-k_{21}+k_{11}-k_{22}$$

$$\gamma=\frac{2}{c}(k_1-k_2)$$



ing. There may, however, be specificity, too, because both proteins do not necessarily recognize both sequences equally well, nor do they recognize unrelated sequences at all. The sequences provide a specific binding site for initiating replication. The differences in binding strength for the four possible interactions of E_1 and E_2 with I_1 and I_2 may be slight. These differences, indicated by the line strengths, however small they are, will have drastic consequences, as follows from an inspection of the corresponding fixed-point diagrams (Table 17). We may distinguish four cases:

(1) E_1 favors I_1 over I_2, and E_2 favors I_2 over I_1 (Fig. 53a).
Consequence: I_1 and I_2 both are hypercyclically enforced by their respective enzymes, leading to strong competition. Only one of the competitors can survive, even if they are selectively equivalent.
(2) E_1 favors I_1 over I_2, and E_2 also favors I_1 over I_2 (Fig. 53b).
Consequence: I_1 will win the competition and I_2 will die out.
(3) E_2 favors I_2 over I_1, and so does E_1 (Fig. 53c).
Consequence: I_2 is now the winner, while I_1 dies out.

(4) E_1 favors I_2 over I_1, and E_2 favors I_1 over I_2 (Fig. 53d).

Consequence: Here we obtain a mutual, hypercyclic stabilization of I_1 and I_2.

It is important to note that small differences suffice to produce the behavior outlined above. In this respect it is of interest to see what happens if both E_1 and E_2 are exactly equivalent in their treatment of I_1 and I_2. Here we have complete impartiality, regardless of how much the population numbers x_i or y_i differ. I_1 or I_2 can die out in consequence of a fluctuation catastrophe, since there is no mutual stabilization present as in case 4. On the other hand, the fluctuations do not amplify themselves, and if the population numbers are large enough, a fluctuation catastrophe will practically never occur. Quite different in this respect is case (1), mentioned above. Only for exactly equal population numbers of I_1, I_2, E_1, and E_2 is here the system in a dynamically balanced state. A small fluctuation may disturb the balance and then, through self-amplification, inevitably will lead to selection of one of the two species. The same is true for any ensemble in which each messenger provides help for its own replicase only (cf. Fig. 47). The coupling resulting from a common translation function – all replicases (utilizing their RNA-recognition sites) may function simultaneously as activating enzymes – is not sufficient to enforce a coexistence. As in the system shown in Figure 45, there will be only one survivor, and translation function will subsequently break down.

The exact criteria for hypercycle formation are derived in Table 17. The figures give a clear representation of the stability ranges in terms of generalized coordinates referring to the rate parameters.

We have thus obtained an evolution principle for hypercycles. This kind of organization can emerge from a single quasi-species distribution, as soon as means of reaction coupling develop. The prerequisites for coexistence of precursors can be met generally only by closely related mutants. Thus the emergence of hypercycles requires the pre-existence of a molecular Darwinian system, but it will then lead to quite new consequences. The evolution principle is effective even with very small differences in rate parameters and hence responds sensitively to small changes brought about by mutations. Given a quasi-species distribution with developing interactions among the constituents, regardless, of how weak these interactions are, a hypercyclic organization will inevitably emerge whenever such interactions arise.

The hypercycle will also grow inevitably by way of mutations toward larger complexity (Fig. 54, 55). The evolution principle can be generalized by induction so as to apply to any n-membered hypercycle. A mutant I′ then may either replace I, die out, or enlarge the hypercycle to a size comprising $n+1$ members (cf. XVI. 10.). More general evolution criteria can be derived as indicated in Table 18.

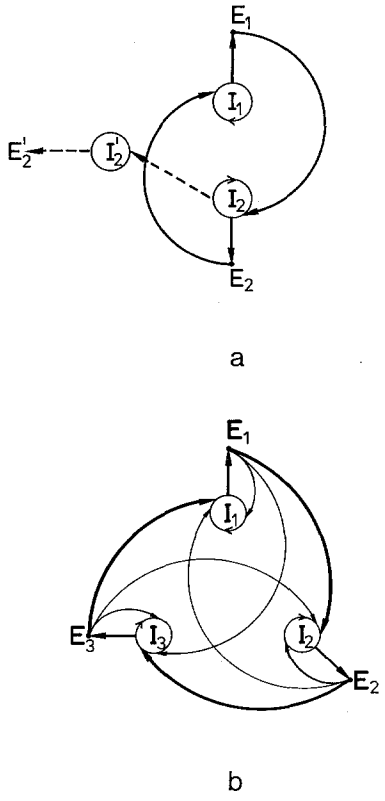


Fig. 54. *The generalization of the evolution principle* of hypercycles is illustrated in this diagram. The couplings have to fulfil the criteria derived in Tables 17 and 18, i.e., mutual enhancement has to prevail over self-enhancement (cf. thick lines). (a) A mutant of I_2 appears (I_2'). (b) The mutant (now I_3) is incorporated in the hypercycle

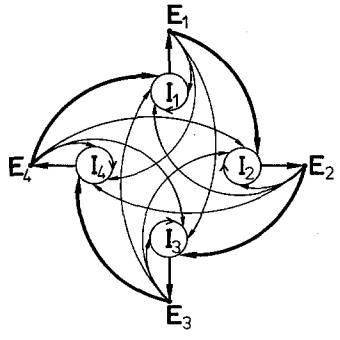


Fig. 55. *The 'realistic' four-membered hypercycle* assigns four messengers I_1 to I_4 (being mutants of a common precursor) to encode for four replicases E_1 to E_4 with common function, but slight preferences in specificity. The minus strands of I_1 to I_4 may concomitantly act as amino acid adapters

Naturwissenschaften 65, 341 – 369 (1978)

Table 18. Generalization of the evolution principle of hypercycles

is explained by the transition of a two-membered to a three-membered system as depicted in Figure 54.
The general rate equations are of the same form as in Table 17. Starting from a stable two-membered hypercycle, introduction of a third member I_3 (e.g., being a mutant of I_2) will frequently change the previously stable fixed point to a saddle point. This is most clearly seen if cyclic symmetry is assumed. Under these conditions, the notation can be simplified to:

$$k_{11}=k_{22}=k_{33}=k_D; \quad k_{13}=k_{21}=k_{32}=k_+; \quad k_{12}=k_{23}=k_{31}=k_-;$$

yielding the following matrix of rate coefficients.

$$\mathbf{K}=\begin{vmatrix} k_D & k_- & k_+ \\ k_+ & k_D & k_- \\ k_- & k_+ & k_D \end{vmatrix}$$

The fixed points and eigenvalues then are:

Corners: $\bar{\mathbf{x}}_1=(c,0,0)$; $\omega_1^{(1)}=(k_- -k_D)c$, $\omega_2^{(1)}=(k_+ -k_D)c$

$\bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3$ analogous

Edges: $\bar{\mathbf{x}}_4=(0, k_D-k_-, k_D-k_+)\dfrac{c}{2k_D-k_+-k_-}$

$$\omega_1^{(4)}=\frac{k_-(k_D-k_-)+k_+(k_D-k_+)+k_+k_- -k_D^2}{2k_D-k_+-k_-}c$$

$$\omega_2^{(4)}=\frac{(k_D-k_+)(k_D-k_-)}{2k_D-k_+-k_-}c$$

$\bar{\mathbf{x}}_5, \bar{\mathbf{x}}_6$ analogous

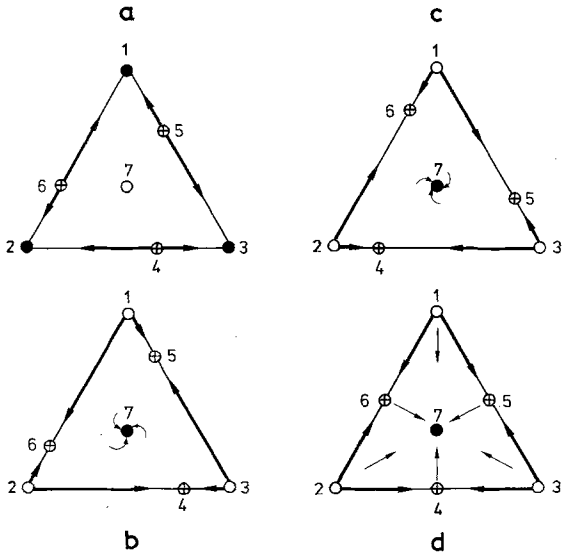
Interior: $\bar{\mathbf{x}}_7=\left(\frac{c}{3},\frac{c}{3},\frac{c}{3}\right)$,

$$\omega_{1,2}^{(7)}=\{2k_D-k_+-k_- \pm i\sqrt{3}(k_+-k_-)\}\frac{c}{6}.$$

Again four cases are of special interest:

a) $k_D>k_+, k_-$
b) $k_+>k_->k_D$
c) $k_->k_+>k_D$
d) $k_+=k_->k_D$

yielding the following fixed-point diagrams:



Large diagonal terms ($k_D \gg k_+, k_-$) lead to competition (diagram a). In the opposite situation, i.e., with large off-diagonal elements of **K**, the three species show cooperative behavior. The sense of rotation around the spiral sink in the center of the simplex is determined by the larger of the two constants k_+ and k_-. No rotational component is observed for equal constants $k_+=k_-$. The central fixed point is then a focus.
The example treated in this table provides a good illustration of the evolution to more complex hypercycles. In the absence of simplifying assumptions concerning the rate constants the analysis becomes quite involved. We refer to a more detailed representation [98], which includes a generalization to arbitrary dimensions.

XVI. Ten Questions

concerning our earliest molecular ancestors and the traces which they have left in the biosynthetic apparatus of present cells.

XVI.1. One RNA precursor?

This question is concerned with the complexity of the first molecules starting any reproducible function. A nucleotide chain of 100 residues corresponds to a complexity of about 10^{60} alternative sequences. If on grounds of stability we restrict ourselves to (AU-doped) GC copolymers only, we are still left with about 10^{30} possible arrangements. In order to achieve one or a few defined sequences, faithful self-reproduction is a necessary prerequisite. It will inevitably lead to Darwinian behavior, with selection of one defined quasi-species. The selected products are determined plainly by an optimal selective efficiency, but their structure depends on their historical route, which is strongly biased by self-copying of smaller oligonucleodtide patterns.

XVI.2. What Does Selective Advantage Mean to a Molecule?

Selective value is defined as an optimal combination of structural stability and efficiency of faithful replication. It can be expressed in quantitative terms related to the physical properties of a molecule in a given environment. Structural stability, resistance toward hydrolysis, and the development of cooperative properties call for elongation. Small oligonucleotides can-

not fold in any stable manner and may therefore easily be hydrolyzed. Furthermore, they do not offer sufficient adhesive strength for faithful copying or for translation. Length, on the other hand, is limited by replication rates and by copying fidelity. The properties of GC-rich sequences have been shown to be advantageous for forming stable copies with extended length. Whether these lengths resemble the sizes of present-day tRNA is uncertain. Sequence homologies have been found in tRNA [74], which indicate some self-copying of internal regions. This, however, may well have happened before codons became assigned. The onset of translation requires strong interactions between adapters and messengers, and these cannot be provided by molecules which are too small.

As soon as translation yields reproducible functions, selective value achieves a new dimension. It must, nevertheless, be expressed, for any given messenger, in terms of structural stability and efficiency of faithful reproduction. These properties now, however, also depend on the qualities (and concentrations) of the translation products. Specific coupling – as required for hypercyclic organization – is hence necessary for any system in which translation products are to be rated for selection and thereby become eligible for evolution. Such coupling is of a catalytic or protective nature.

XVI.3. Why Hypercyclic Organization of Single Mutant Genes Rather than One Steadily Growing Genome?

The answer to this question has been largely given in Part A. For a very primitive translation apparatus an amount of information would be required that corresponds to (or even exceeds) that of present RNA phages. The information of the phage genome can be preserved only with the help of a phage-specific enzyme complex, the availability of which is based on the efficiency of a complete translation machinery, provided by the host cell. If we accept the answers given to the first and second questions, the information needed to start translation must arise from cooperation among several mutants coexisting in the quasi-species distribution, rather than from a mere elongation of one sequence for which primarily no selection pressure would exist.

The hypercyclic stabilization of several coexisting mutants is equivalent to evolution by gene duplication. Originally, mutants appeared as single strands rather than as covalently linked duplicates. Fidelity restrictions would not allow for such an extension of length. Moreover, the probability of obtaining the required mutant combinations in one strand is very low. Sequences consisting of 100 G and C residues have

100 one-error mutants,
4950 two-error mutants
161700 three-error mutants, etc., or

$$N_k = \binom{100}{k}\ k\text{-error mutants}$$

The number of strands containing n mutant genes, each differing from the other in k specified positions (which may be necessary in order to qualify for a function) amounts to

$$\binom{N_k+n-1}{n} \approx \frac{N_k^n}{n!} \quad (\text{for } n \ll N_k),$$

e.g., for $n=4$ and $k=3$ to 3×10^{19} alternative sequences. Given even these small deviations in the multiplied genes, the chance of finding a copy with a favorable combination within one giant strand is almost nil for any reasonably sized population. Each of the *isolated* mutant genes containing three substitutions, however, would be abundantly present in any macroscopic population.

Last but not least, the tRNAs being the adapters for translation must have been present anyway as separate strands. Evolution of a unified genome would have required complicated transcription control right at the start.

The isolated RNA strands, on the other hand, have a natural origin in the quasi-species distribution. All sequences were similar and so must have been their translation products. Whenever one translation product provides coupling functions, all of them will do so, owing to their similarities. Cyclic coupling – as required for hypercyclic organization – may then occur as well. We might even say that hypercyclic organization is most naturally associated with any realistic primitive translation model.

Does the present genome organization, established in prokaryotic cells, offer any clue as to its early structure? Present genes are certainly much larger than the early messengers. Gene elongation, as well as duplication, provided an advantage whenever the steadily increasing fidelity of the enzymic machinery allowed for it. The translation products could gain in sophistication, and more complex multienzyme mechanims could evolve, utilizing differentiated enzymes which had descended from a common precursor. Recombinant mechanisms as utilized by present-day cells will not have been available in primitive systems. The present structure of the prokaryotic genome therefore may have been achieved through elongation of isolated genes, their duplication and triplication to operons and their final mapping onto DNA, which can utilize more advanced means of reproduction so

Naturwissenschaften 65, 341 – 369 (1978)

as to allow for the formation of a unified genome. The present operon sizes correspond well to those which can be handled by a sophisticated RNA replicase (e.g., 1000 to 10000 nucleotides).

XVI.4. Are tRNAs Necessary to Start With?

This question may be alternatively posed as: Why not small oligonucleotide adapters?
Adapters without messengers do not make much sense. Short nucleotide sequences do not qualify as messengers. Decapeptides are already equivalent to almost half a tRNA molecule. Furthermore, short oligonucleotides may be unstable since they lack any tertiary structure. The simplest symmetric structure, i.e., a single loop stabilized by four or five base pairs, requires as many as fifteen nucleotides. Enzyme-free specific recognition of an amino acid involving simultaneously the anticodon loop and the 3′-end of the adapter is possible only with more extended structures. The same is true for interactions between two adjacent adapters, necessary for the stabilization of the messenger-peptidyl-tRNA complex, or for the conformational change (e.g., from HF to fh) which may facilitate the transport of the growing peptide chain along the messenger. G. Maass et al. [99] recently reported such a conformational change in the anticodon loop of tRNA, which they recorded by observing a fluorescence change of the Y base. The effect appeared to be absent in the anticodon-loop fragment (i.e., a decanucleotide having the sequence of the anticodon loop). All of this suggests that insufficiently long RNA sequences do not qualify for adapter function.
We may then ask: What distinguishes an adapter from a messenger? They both require comparable minimum lengths. They both have to be specifically folded structures, through which they may become reproducibly recognizable by coupling factors.
Since each tRNA *and* each messenger require a coupling factor, e.g., a replicase, favoring their selective stabilization, dual functions of the RNA sequences are indispensable. Plus and minus strands of a given RNA sequence may thus be utilized jointly as messenger and adapter.

XVI.5. Do Present-day tRNAs Provide Clues about their Origin?

Similarities in structure might either be the consequence of adaptation to a common goal or, alternatively, indicate a common ancestor. Present tRNAs show many points of correspondence [75] in their structures. Are we able to infer a common ancestor from these analogies? According to an analysis carried out by T.H. Jukes [76], this question may be answered with a cautious 'yes'. Why one must be cautious may be illustrated with an example. One of the common features exhibited by all prokaryotic and eukaryotic tRNAs studied so far is the sequence TΨCG in the so-called T-loop, a common recognition site for ribosomal control. Recent studies of methanogenic bacteria [77] revealed that these microorganisms, which are thought to be the 'most ancient divergences yet encountered in the bacterial line', lack this common feature of tRNA, but rather contain a sequence ΨΨCG in one and UΨCG in another group. Although this finding does not call in question but rather underlines the close evolutionary relations of this class of microorganisms with other prokaryotes, it shows definitely that whole classes may concordantly adopt commont features. This is especially true for those molecules that are produced by a common machinery, such as the ribosome, which is the site of synthesis of all protein molecules.
Figure 56 shows an alignment of the sequences of four tRNAs from *E. coli*, which we think are the present representatives of early codon adpaters. Unfortunately, the sequence of the alanine-specific tRNA adapted to the codon GCC was not available. If we compare this species, which has the anticodon ^{5A}UGC, with its correspondent for valine, which has the anticodon ^{5A}UAC, we observe a better agreement

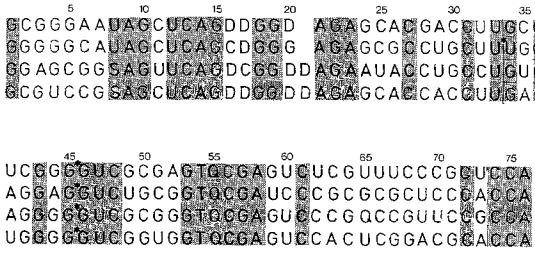


Fig. 56. *Alignment of the sequences of tRNAs* for the amino acids gly, ala, asp, and val. Unfortunately, the sequence referring to the codon GCC (for ala) is not yet available. Correspondences between gly- and ala-tRNAs are supposed to be closer for the correct sequence referring to the anticodon GCC (as suggested by the similarities between the two sequences for ala and val, referring to the anticodons *UGC and *UAC, resp.). The sequences show that base-paired regions consist predominantly of GC, and that close correspondences indicate the kinship between gly/ala and asp/val (cf. S in position 8 for asp and val instead of U for gly and ala, or the insertion of D between position 20 and 21 for asp and val). A = adenosine, *A = 2MA = N(2)-methyladenosine, C = cytidine, D = 5,6-dihydrouridine, G = guanosine, *G = 7MG = N(7)-methylguanosine, Q = Ψ = pseudouridine, S = thiouridine, T = ribosylthymine, U = uridine, *U = 5AU = 5-oxyacetyluridine

with the latter than with the one listed in the alignment (57 vis-à-vis 54 identical positions). Hence the correct alanine-tRNA with the anticodon GGC may have more coincidences with the gly-tRNA listed than for the 44 positions shown. Apart from this 'corporal defect' the data reveal

1. That all representatives agree in more than half of the positions (33 including the 'wrong' ala or 41 for gly, asp, and val),
2. That the subgroup gly/ala is distinguished from the subgroup asp/val by several features (thio-uridine 'S' instead of U in position 8, insertion of a 5,6-dihydro-U between position 20 and 21),
3. That all representatives have a pronounced excess of G and C over the A and U residues (or their derivatives), especially in the base-paired regions.

A comparison with other tRNA sequences, furthermore, indicates that these features—although certainly not uncommon among most tRNAs—are especially pronounced for this group. In particular, correspondences are as close as for different adapters of the same amino acid in the same organism.

One finding is particularly illuminating. If we compare the sequences of two adapters with complementary anticodons (e.g., asp and val) the coincidences between both plus strands of the tRNAs are much more pronounced than those between one plus strand (read from 3′ to 5′) and the other minus strand (read from 5′ to 3′). Actually, if we compare in this way the plus and minus strands of the same tRNA, the agreement is better. These coincidences are only the expression of the remarkable internal symmetry of tRNA, which places the anticodon almost exactly in the middle of the sequence and thereby allows for the formation of a symmetric two-dimensional pattern. We may rate this property as an indication of the early appearance of tRNA as an independent replicative unit. The requirement for the plus and minus strands to assume a similar pattern is important only if these represent independent replicative units rather than being structurally integrated into a large sequence of a genome, as they appear to be nowadays. We find a similar effect for phage RNAs or their variants, which have to multiply as single replicative units [78].

On the other hand, adaptation of tRNA to a common machinery must have brought about common deviations from symmetries originally required. The fact that the mirror images for plus and minus strands of the same tRNA show still more symmetric resemblance than those for plus and minus strands of tRNAs with complementary anticodons suggests that both tRNAs evolved as mutants of the same rather than of two complementary strands. We may then conclude that the present adapters for the codons GGC (gly), GCC (ala), GAC (asp), and GUC (val) derived from one quasi-species as single error mutants of a common ancestor. However, the original symmetry was not sufficient (why should it have been?) to allow adapter functions to derive from both the plus *and* the minus strand of a given RNA.

XVI.6. How Could Comma-free Messenger Patterns Arise?

The first messengers must have been identical with the first adapters (or their complementary strands). There is, indeed, a structural congruence behind adapter and messenger function. Whatever codon pattern occurs in the messenger sequence, it must have its complementary representation at the adapter. In primitive systems such a requirement could be most easily met by utilization of a common structural pattern for both types of molecules, such that the first adapters are the minus strands of the first messengers (if we define the plus strand as always being associated with a message) and that certain symmetries of structure allow both the plus and the minus strand to be recognized by the same replicase.

The first extended RNA molecules were rich in G and C, a consequence of selection based on criteria of structural stability and fidelity of copying. Molecules with a common codon pattern, such as GGC/GCC, require primer instruction (e.g., via catalysts or via exposed loops of RNAs present) with subsequent internal duplication. This will inevitably lead to structures that contain at least two codon patterns with internal complementarity, e.g., 5′GGC3′ and 3′CCG5′.

There is a good example for the efficiency of internal pattern duplication in the de-novo synthesis and amplification of RNA sequences by phage replicases. If Qβ replicase is severely deprived from any template, it starts to 'knit' its own primers, which it then duplicates and amplifies (selectively) until finally a uniform macroscopic population of RNA sequences—a few hundred nucleotides in length—appears. Under different environmental conditions, different (but uniform) sequence distributions are obtained [8]. S. Spiegelman, D. Mills and their co-workers have sequenced some of these 'midivariants,' all of which contain the specific recognition site for Qβ replicase [78]. Further experiments [73] have thrown light on the mechanism of this de-novo synthesis, showing that small pieces corresponding to sequences that are recognized by the enzyme are made as primers and then internally duplicated and selectively amplified. Earlier studies [22] have shown that, in particular, the sequences CCC(C) and UUCG can be recognized

Naturwissenschaften 65, 341–369 (1978)

Fig. 57. *Alignment of the sequence of Qβ-midivariant* (determined by S. Spiegelman et al. [78]) with an artificial sequence composed of CCC(C)- and UUCG-blocks, as well as their complements [GGG(G) and CGAA]. Agreement at 169 of 218 positions suggests that midivariant is a de-novo product made by the enzyme Qβ-replicase, which possesses recognition sites for CCC(C) and UUCG (EF Tu). The kinetics of de-novo synthesis indicates a tetramer formation at the enzymic recognition sites, followed by some internal self-copying with occasional substitutions. The specific midivariant usually wins the competition among all appearing sequences and hence seems to be the most efficient template. The process demonstrates how uniform patterns can arise in primitive copying mechanisms

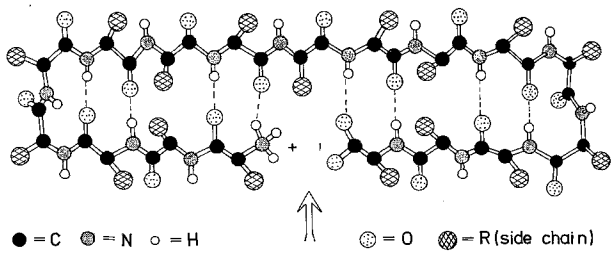


Fig. 58. *A simple enzyme precursor* is represented by a β-folded structure of some 15 to 25 amino acids (requiring messengers of 45 to 75 nucleotides). The active site includes a terminal amino group that is a very efficient proton donor (pK ~ 8), a terminal carboxylic group that acts as proton acceptor, and a catalytically active side chain (e.g., asp or ser). Many alternatives could be designed, only some of which have the correct pitch of the twisted chains to yield an efficient active site

by the enzyme. UUCG corresponds to the sequence TΨCG common to all tRNAs and known to interact specifically with the ribosomal elongation factor EF Tu, which acts as a subunit in the Qβ-replicase complex. An alignment of the midivariant with a sequence made up solely of the two oligonucleotides mentioned and their complementary segments – GGG(G) and CGAA – shows agreement in more than three-quarters of the positions, indicating the efficiency of internal copying of primer sequences (Fig. 57).

In a similar way we may think of the existence of primordial mechanisms of uniform pattern production. If among the many possible patterns 5′GGC/5′GCC and possibly also 5′GAC/5′GUC appeared, those messenger patterns could have started a reproducible translation according to the mechanism of Crick et al. [3] and have been capable of selective amplification with the help of their reproducible translation products.

XVI.7. What Did the First Functionally Active Proteins Look Like?

The simplest protein could be a homogeneous polypeptide, e.g., polyglycine. Does it offer any possible catalytic activity? This is a question that can and should be answered with experiments. With mixed sequences, including a sufficiently large number of residues, say about fifteen to thirty, β-sheet structures may form with an active center, in which the terminal carboxylic group is placed in a defined position near the terminal amino group (Fig. 58). The proximal distance varies with the chain length, since the β-structure involves a twist among both antiparallel chains [79]. The pK of the terminal amino group is around eight, hence the catalytic site contains at least an efficient proton donor-acceptor system. Alternating gly-ala residues are very favorable for the formation of β-structures. However, there are serious solubility problems for chains consisting exclusively of gly and ala, which would restrict them to interfaces only.

The folding of β-sheets has been studied by P.Y. Chou and G.D. Fasman [80], who analyzed X-ray data for 29 proteins in order to elucidate 459 β-turns in regions of chain reversal. The three residues with the highest β-turn potential in all four positions of the bend include gly and asp, while in regions following the β-turn, hydrophobic residues are predominant.

An important prerequisite of catalytic efficiency is the defined spatial arrangement of the terminal groups. The utilization of two or more classes of amino acids may be necessary for stabilizing a reproducible folding. β-Sheets have long been known to be important building elements of protein structure. According to M. Levitt [81], they may be utilized in a very general manner to stabilize active conformations of proteins.

The large abundance of glycine and alanine might have determined in essence the appearance of the first proteins, but polar side chains are indispensible for the solubility of longer sequences. Four amino acid classes would of course offer much more flexibility. If aspartic acid and valine were the next two

candidates, globular structures might have formed, stabilized by hydrophobic interactions of the side chains of valine and alanine and solubilized by the carboxylic side chains of aspartic acid. This residue further offers many possibilities for forming specific catalytic sites with the participation of divalent metal ions.

Our imagination is taxed to estimate the vast number of possibilities. Experiments that are supposed to test various structures with respect to their efficiency in discriminating between RNA sequences and their structural features are under way. Results obtained with ribonucleases [82] encourage one to seek a 'minimum structure', able to recognize RNA sequences specifically.

XVI.8. Are Synthetases Necessary to Start With?

In the three-dimensional structure of present-day tRNAs (cf. Part A, Fig. 14) the anticodon loop is fixed at a considerable distance from the amino acyl site. Such a structure is adapted to the functional needs of present tRNA molecules, imposed by the ribosomal mechanism and by the structure of synthetases. On the other hand, it is known that tRNA can undergo conformational changes that drastically alter its shape and dimensions. R. Rigler and his coworkers [83] studied conformational lifetimes as well as rotational relaxation times by fluorescence methods and concluded the existence of at least three different rapidly interconverting conformational states. Analogous results were obtained by T. Olson et al. [84], who used laser light-scattering techniques. The population of the different conformational states depends strongly on magnesium-ion concentration. It is important, again, to note that under conditions that correspond to those present in sea water (Mg^{2+}; ~ 50 mM), a conformer is present that differs in shape from the L-form found by crystallographic studies, being considerably more cylindric.

This point is stressed because it is most relevant to the question raised. Early enzymes were made of only a very limited number of amino acid residues and therefore cannot have been very bulky globular structures. In order to guarantee a unique assignment of an amino acid to an anticodon, either enzymes as sophisticated as present-day aminoacyl synthetases had to be available, or else the tRNA structure had to allow a much closer contact between the aminoacyl and anticodon sites than the L-form does, in order to admit a simultaneous checking of both sites. The high mutation rate at early stages would otherwise very soon have destroyed any unique coincidental correspondence between these two sites. On the other hand, the conformational transition is still required since the mechanism of peptide-bond formation (cf. Fig. 48) calls for a well-defined separation of the messenger and the growing peptide chain. The data quoted invite reflection about such possibilities. If, on the other hand, a structure similar to the pattern c) shown in Figure 49 is likely to arise, the first amino-acid assignments might even have been made without enzymic help. The tRNA structure as such certainly offers sufficient subtlety for specific recognition. It has been noted [85] that the fourth base from the 3′-end (i.e., the one following 3′ACC) is somehow related to the anticodon. The primary expectations regarding a unique correlation for all tRNAs finally did not materialize. However, such a correlation may have played a key role in the early specific recognition of amino acids by tRNAs. Referring to data from *E. coli* and T_4 phage, the nucleotides in the position following 3′ACC are: U for gly, A for ala, G for asp, and A for val. It was certainly important for early adapters to ensure unique assignment by sufficiently discriminative sites. This property might have been partially lost during later phases of evolution. This is admittedly a speculation and calls for experimental confirmation.

To conclude: Synthetases may have been dispensible at the very early stages, but tRNAs finally turned out to be an unsatisfactory attempt by Nature to make enzymes from nucleic acids. More efficient recognition may have evolved from the coupling factors, which were predestined to recognize tRNA-like structures specifically.

XVI.9. Which Were the First Enzymes?

If synthetases are not really necessary for a *start* of translation (and this is a big 'if!'), we are left with the coupling factors, probably replicases, as the only absolute primary requirements for a coherent evolution of translation. Via such a function, a selective advantage occurring in a translation product can be most efficiently fed back onto the messenger. Hence specific replicases (all belonging to one class of similar protein molecules) not only provide the prerequisites for hypercyclic coupling, but also turn out to be most important for the further evolution of proteins, since only they can tell the messenger what is phenotypically advantageous and how to select for it at the genotypic level, i.e., by enhanced synthesis of the particular messenger. As will be seen in the next paragraph, such a selective coupling between geno- and phenotypic levels works best in combination with spatial separation or compartmentation.

Naturwissenschaften 65, 341–369 (1978)

Next, of course, we have to look for catalytic support for the various translation functions. If replicases have established a defined relationship with tRNA-like messengers (including both the plus *and* the minus strands), their recognition properties may well be utilized for synthetase and translatase (i.e., preribosomal) functions. In other words, a gene duplicate of a replicase may well be the precursor of a synthetase messenger as well as of a translation factor such as EF Tu, the more so since the chemistry of replicase and transfer function is very similar and in present systems appears to be effected by similar residues.

Dual functions with gradual divergence may have been a very early requisite of replication and translation mechanisms, just as gene duplication was one of the main vehicles of evolution at later stages.

Those dual functions have clearly left their traces in present cell organelles, and viruses have utilized them as well for their postbiotic evolution in the host cell. The genome of the phage Qβ encodes for only one subunit of its replicase, but utilizes three more factors of the host cell, which have been identified as the ribosomal protein S_1 and the elongation factors EF Tu and EF Ts [87, 88].

Ch. Biebricher [89] has studied the properties of these factors and found that they are involved simultaneously in several functions of ribosomal control, utilizing their acquired property, namely, to recognize tRNA molecules. He argues that also the β-factor of Qβ replicase, which is encoded by the phage genome, has its precursor in the *E. coli* cell, and this seems, indeed, to be the case. Using immunologic techniques, he was able to identify a protein containing EF Tu and EF Ts that behaves like a precursor of the Qβ replicase in uinfected *E. coli* and that appears to be involved in an – as yet unspecified – RNA-synthesis function of *E. coli* [87]. Further, similar correspondences may yet be found with synthetases. It seems that once a certain function has been developed – such as the ability to recognize certain types of RNA – then Nature utilizes this function wherever else it is needed (e.g., specific replication, ribosomal transport and control, amino acid activation).

In some respects the formation of RNA phages may well have mimicked the evolution of early RNA messengers. Phages utilize as many host cell functions as possible except one, namely, specific recognition of their own genome (i.e., coupling via specific replication). Different phages (e.g., Qβ, Ms2, R17) inherit different recognition factors [9], although they all derive from a common ancestor in the host cell. In Part A it was also shown that the primary phase of RNA-phage infection is equvalent to a simple hypercyclic amplification process.

XVI.10. Why Finally Cells with Unified Genomes?

Hypercycles offer advantages for enlarging the information content by functional integration of a messenger system, in which the single replicative unit is limited in length due to a finite fidelity of copying. The increase in information content allows the build-up of a reproducible replication and translation apparatus, by which the translation products can evolve to higher efficiency. This will allow better fidelities, which in turn will increase the information content of each single replicative unit and thereby, again, enhance the quality of the enzymes. Simultaneously, as shown in Section XV, the hypercycle itself will evolve to higher complexity by integrating more differentiated mutant genes.

Increase of information content will not only produce better enzymes; it may also allow each replicative unit to inherit information for more than one enzyme. Dual functions can thereby be removed from the list of earlier evolutionary constraints, i.e., duplicated messengers may develop independently, according to the particular functional needs of their translation products. This may have been the origin of operon structures with control mechanisms for simultaneous replication of several structural genes. Replicases may thus have evolved to common polymerases associated with specific control factors for induction or repression.

After having realized the advantages of functional coupling, which seems to be a requirement for any start of translation, we should ask why functional coupling has finally been replaced by complete structural integration of all genes, the genomes of even the most primitive known cells being structural units. So where are the limitations of hypercyclic organization and what improvements can be made in it?

In a system controlled by functional links we have to distinguish two kinds of mutations. One class will primarily change the phenotypic properties of the messenger itself and thereby alter its target function with respect to a specific replicase or control factor. These mutations are especially important in the early phases of evolution owing to the important role of phenotypic properties of RNA structures. Those target mutations will immediately become selectively effective, advantageous mutations will be fixed, and disadvantageous ones will be dismissed.

The second kind of mutation – which may or may not be neutral with respect to the target function – refers to phenotypic changes in the translation products. The more precisely specified the messengers are, the more specifically a mutation may alter the function of the translation product.

Whether or not a mutant is specifically favored by

selection depends only on the target function, regardless of whether the translation product is altered in a favorable or an unfavorable sense or whether it remains neutral. For the later stages of precellular evolution the most common consequence of a mutation will be a phenotypic change in the translation product coupled with an unaffected target function. The mutant may then proliferate further, but it is not specifically selected against its former wild type, nor would the system select against the mutant, if its translation product proves to be unfavorable. What should really be achieved is a rating of the system as a whole. This may be accomplished by spatial separation of the messenger systems, by niches or – even more efficiently – by compartmentation. A messenger in a given compartment can enrich its environment with its own translation products and compete with other compartments using its efficiency of proliferation. To a limited extent this is also possible simply by spatial separation. However, a compartment without hypercyclic organization does not work at all. The enhanced competition among all messengers in the limited living space of the compartment would destroy any cooperative function.

A compartment could proliferate more efficiently by correlating its own reproduction with the re-duplication of its total gene content. This, of course, requires a fairly involved control mechanism, which could be facilitated by the integration of all genes into one giant replicative unit. Such an individualization of the total compartment requires high fidelity in the replication machinery. In Part A we compared the information content of various stages of life with their corresponding (and observed) replication fidelities (cf. Table 4).

The individualization of compartments is probably connected with the transition from RNA genes or operons to DNA genomes, since only the mechanism of DNA replication could guarantee sufficiently high fidelity. The new individualized unit was the integrated Proto-cell. The previous functional organization of genes and gene products has been superseded and amended by a coupled structural and functional organization. A closer study of the cyclic arrangement of genetic maps may still reveal some remnants of the origins of structural organization, although recombinative epigenetic effects may have covered many of the traces.

As a consequence of unification and individualization, the net growth of (asexual) multiplication of cells obeys a first-order autocatalytic law (in the absence of inhibitory effects). The Darwinian properties of such systems allow for selective evolution as well as for coexistence of a large variety of differentiated species. The integrated unit of the cell turns out to be superior to the more conservative form of hypercyclic organization.

On the other hand, the subsequent evolution of multicellular [90] organisms may again have utilized analogous or alternative forms of hypercyclic organization (nonlinear networks) applied to cells as the new subunits, and thereby have resembled in some respect the process of molecular self-organization.

XVII. Realistic Boundary Conditions

A discussion of the 'realistic hypercycle' would be incomplete without a digression on realistic boundary conditions. We shall be brief, not because we disregard their importance in the historical process of evolution – the occurrence of life on our planet is after all a historical event – but because we are aware of how little we really can say. While the *early stages of life,* owing to evolutionary coherence, have left at least some traces in present organisms, there are no corresponding remnants of the *early environment.*

In our discussion so far we have done perhaps some injustice to experiments simulating primordial, template-free protein synthesis, which were carried out by S.W. Fox [91] and others (cf. the review by K. Dose and H. Rauchfuss [92]). It was the goal of our studies to understand the early forms of organization that allowed self-reproduction, selection, and evolutionary adaptation of the biosynthetic machinery, such as we encounter today in living cells. Proteins do not inherit the basic physical prerequisites for such an adaptive self-organization, at least not in any obvious manner as nucleic acids do. On the other hand, they do inherit a tremendous functional capacity, in which they are by far superior to the nucleic acids. Since proteins can form much more easily under primordial conditions, the presence of a large amount of various catalytic materials must have been an essential environmental quality. Research in this field has clearly demonstrated that quite efficient protein catalysis can be present under primordial conditions.

Interfaces deserve special recognition in this respect. If covered with catalytically active material they may have served as the most favorable sites of primordial synthesis. The restriction of molecular motion to the dimensions of a plane increases enormously the efficiency of encounters, especially if sequences of high-order reactions are involved.

L. Onsager [93] has emphasized that under primordial conditions the oceans must have been extensively covered with layers of deposited hydrophobic and hydrophilic material cf. also [94]). Those multilayers must have offered favorable conditions for a primordial preparative chemistry. In view of the obvious

Naturwissenschaften 65, 341 – 369 (1978)

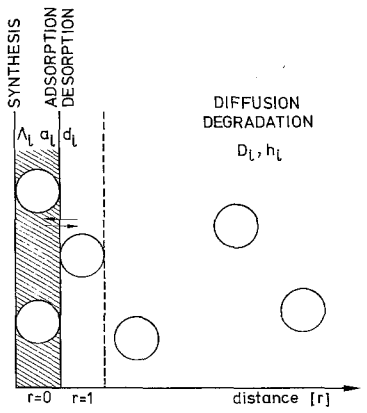


Fig. 59. Schematic representation of a heterogeneous reaction model including hypercyclic coupling. Three spatial regions are distinguished: $r=0$ bound to interface, $r=1$ transition layer at interface, $r>1$ bulk of solution phase. Diffusion to and from interface is superimposed on chemical reactions proceeding according to a hypercyclic scheme

advantages offered by interfaces we have examined the properties of hypercycles under corresponding environmental boundary conditions.

As a simple model we consider a system such as that depicted schematically in Figure 59. Polymer synthesis is restricted to a surface layer only ($r=0$), which has a finite binding capacity for templates and enzymes. The kinetic equations are similar to those applying to homogeneous solutions except that we have to account explicitly for diffusion. We distinguish a growth function that refers to the surface concentrations of replicative molecules and enzymes. Diffusion within the surface is assumed to be fast and not rate-determining. Adsorption and desorption of macromolecules is treated as an exchange reaction between the surface layer ($r=0$) and a solution layer next to the surface ($0<r\leqq 1$). Decomposition may occur at the interface and/or (only) in the bulk of the solution. Finally, transport to and from the interface is represented by a diffusion term.

Depending on the mechanism of synthesis assumed, it may be necessary to consider independent binding sites for both templates and enzymes. We used this model to obtain some clues about the behavior of hypercycles with translation (cf. Sect. IX in Part B). Numerical integration for several sets of rate parameters was performed according to a method described in the literature [95]. Three characteristic results – two of which are in complete analogy to the behavior of hypercycles in homogeneous solutions – can be distinguished:

(A) At very low concentrations of polynucleotides and polypeptides or large values of K_i [see Eqs. (73), (75), and (79) in Part B], the surface densities of polymers do not approach a steady-state value but decrease either monotonically or in damped oscillations. Consequently, the macromolecules die out after some time (Fig. 60).

(B) Above a certain threshold value of total concentration, we find limit cycle behavior in systems with $n<4$. The situation is analogous to the low-concentration limit in a homogeneous solution (Fig. 61).

(C) At sufficiently high concentrations we finally obtain a stationary state:

$$\lim_{t\to\infty}\frac{\partial x_i}{\partial t}=0, \qquad \lim_{t\to\infty}\frac{\partial y_i}{\partial t}=0$$

and $\bar{x}_i>0$, $\bar{y}_i>0$, $i=1,2\ldots n$ (Fig. 62), x_i and y_i being the concentrations of enzymes and messengers, respectively, $\bar{x}_i$ and $\bar{y}_i$ their final stationary values, and t the time.

In systems of lower dimensions ($n\leqq 4$) behavior of types (A) and (C) only was observed.

These model calculations were supplemented by several studies of closely related problems using stochastic computer-simulation techniques. The results again showed the close analogy of behavior of hypercycles at interfaces and in homogeneous solution (as described in detail in Part B).

Consideration of realistic boundary conditions is a point particularly stressed in papers by H. Kuhn [96]. We do not disagree with the assumption of a 'structured environment', nor do we know whether we can agree with the postulation of a very particular environment, unless experimental evidence can be presented that shows at least the usefulness of such postulates.

Our models are by no means confined to spatial uniformity (cf. the above calculations). In fact, the logical inferences behind the various models – namely, the existence of a vast number of structural alternatives requiring natural selection, the limitation of the information content of single replicative units due to restricted fidelities, or the need for functional coupling in order to allow the coherent evolution of a complete ensemble – apply to *any* realistic environment. Kuhn's conclusion that the kind of organization proposed is 'restricted to the particular case of spatial uniformity' is beside the point. Who would claim today, that life could *only* originate in porous material, or at interfaces, or within multilayers at the surface of oceans, or in the bulk of sea water? The models show that it may originate – with greater or lesser likelihood – under any of those boundary conditions, if – and only if – certain criteria are fulfilled. These criteria refer to the problem of generation and accumulation of information and do not differ qualitatively when different boundary conditions are applied.

Much the same can be said with respect to temporal uniformity or nonuniformity. It has been shown in Part B that selection criteria may assume an especially simple form if they apply to steady-state conditions. Since they refer to relative rather than to absolute reaction rates, they are qualitatively the same, regardless of whether the system is growing, oscillating, or in a stationary state.

It is true that annealing is a useful procedure for many problems related to phase separations. Whether, however, thermal fluctuations serve equally well for selection of longer polynucleotides, remains to be shown by experiments.

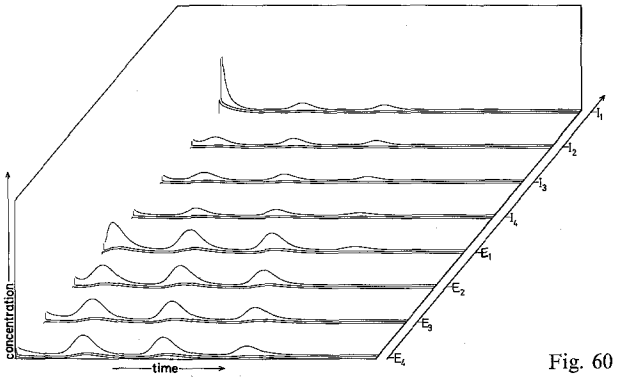


Fig. 60

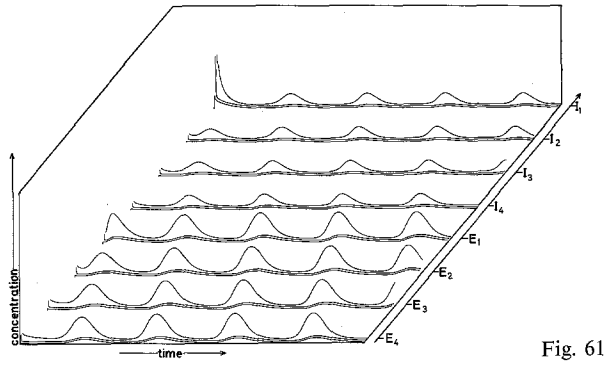


Fig. 61

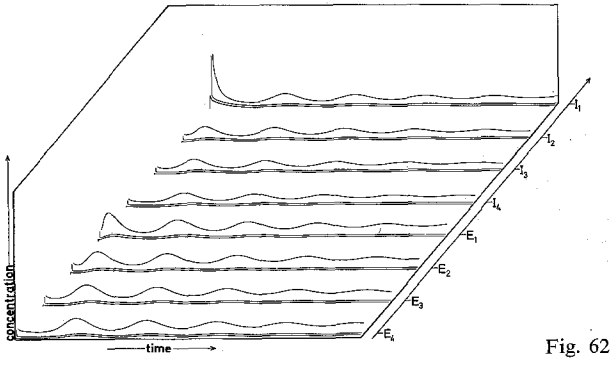


Fig. 62

In order to decide whether fluctuations of temperature improve the selection of strands with higher information content, one must analyze carefully the relative temperature coefficients of all processes involved. The tempererature coefficient of hydrolysis is likely to be the largest of all. *Instructed* replication is by no means generally enhanced at high temperatures. The incoming nucleotide has to bind cooperatively to its complementary base at the template, at the same time utilizing the stacking interaction with the top bases in the growing chain. This is not possible above the melting point of the templates. These considerations apply to any kind of environment, be it an aqueous bulk phase, a surface layer, or a compartment in a coarse-grained or porous material.

The important point for raising the information content is the relative strength of complementary vis-à-vis noncomplementary interactions. Discrimination generally works better at lower than at higher temperatures. S. Miller and L. Orgel ([63], p. 126) conclude from their experimental data:

"We do not know what the temperature was in the primitive ocean, but we can say that the instability of various organic compounds and polymers makes a compelling argument that life could not have arisen in the ocean unless the temperature was below 25° C. A temperature of 0° C would have helped greatly, and −21° C would have been even better. At such low temperatures, most of the water on the primitive earth would have been in the form of ice, with liquid sea water confined to the equatorial oceans.

There is another reason for believing that life evolved

Figs. 60 to 62. Solution curves, obtained by numerical integration, for a system of partial differential equations corresponding to the model depicted in Figure 59. The rate equations account for a growth function A_i as introduced in Part B, which refers to a four-membered hypercycle with translation (B. IX) and which has nonzero catalytic-rate terms only at the interface ($r=0$). The equations furthermore take care of adsorption and desorption (a_i, d_i describing the transition of particles between $r=0$ and $r=1$), hydrolysis (effective at $r \geqq 1$), and diffusion in the bulk of solution ($r>1$, i.e., to and from the transition layer at $r=1$). Each set of three curves refers to the three spatial positions $r=0$ (upper), $r=1$ (medium) and $r=2$ (lower). Figures 60 to 62 differ only in the assumption of different values for the stability constants of the catalytically active complexes $I_i \times E_{i-1}$, which are highest (0.16) in Figure 60, intermediate (0.06) in Figure 61, and lowest (0.04) in Figure 62. The balance between production and removal is sufficient to make the assumption of a dilution flux dispensable. As a consequence of the values chosen (relative to the uniform parameters f_i, k_i – according to B,IX – a_i, d_i and D) autocatalytic production cannot compete with removal by transport and decomposition in Figure 60, where all partners I_i and E_i die out. In both other cases a stable hypercyclic organization is established at the interface, where population numbers are either oscillatory (Fig. 61) or stationary (Fig. 62)

at low temperatures, whether in the oceans or lakes. All of the template-directed reactions that must have led to the emergence of biological organization take place only below the melting temperature of the appropriate organized polynucleotide structure. These temperatures range from 0° C, or lower, to perhaps 35° C, in the case of polynucleotide-mononucleotide helices.

The environment in which life arose is frequently referred to as a warm, dilute soup of organic compounds. We believe that a cold, concentrated soup would have provided a better environment for the origins of life."

XVIII. Continuity of Evolution

It has been the object of this final part of the trilogy to demonstrate that hypercycles may indeed represent realistic systems of matter rather than merely imaginary products of our mind.

Evolution is conservative and therefore appears to be an almost continuous process – apart from occasional drastic changes. Selection is in fact based on instabilities brought about by the appearance of advantageous mutants that cause formerly stable distributions to break down. The descendents, however, are usually so closely related to their immediate ancestors that changes emerge very gradually. Prebiotic evolution presents no exception to the rule.

Let us summarize briefly what we think are the essential stages in the transition from the nonliving to the living (cf. Fig. 63).

1. The first appearance of macromolecules is dictated by their structural stability as well as by the chemical abundances of their constituents. In the early phase, there must have been many undetermined protein-like substances and much fewer RNA-like polymers. The RNA-like polymers, however, inherit physically the property of reproducing themselves, and this is a necessary prerequisite for systematic evolution.

2. The composition of the first polynucleotides is also dictated by chemical abundance. Early nucleic acids are anything but a homogeneous class of macromolecules, including L- and D-compounds as well as various ester linkages, predominantly 2′–5′ besides 3′–5′. Reproducibility of sequences depends on faithfulness of copying. GC-rich compounds can form the longest reproducible sequences. On the other hand, AU substitutions are also necessary. They cause a certain structural flexibility that favors fast reproduction. Reproducible sequences form a quasi-species distribution, which exhibits Darwinian behavior.

3. Comma-free patterns in the quasi-species distribution qualify as messengers, while strands with exposed

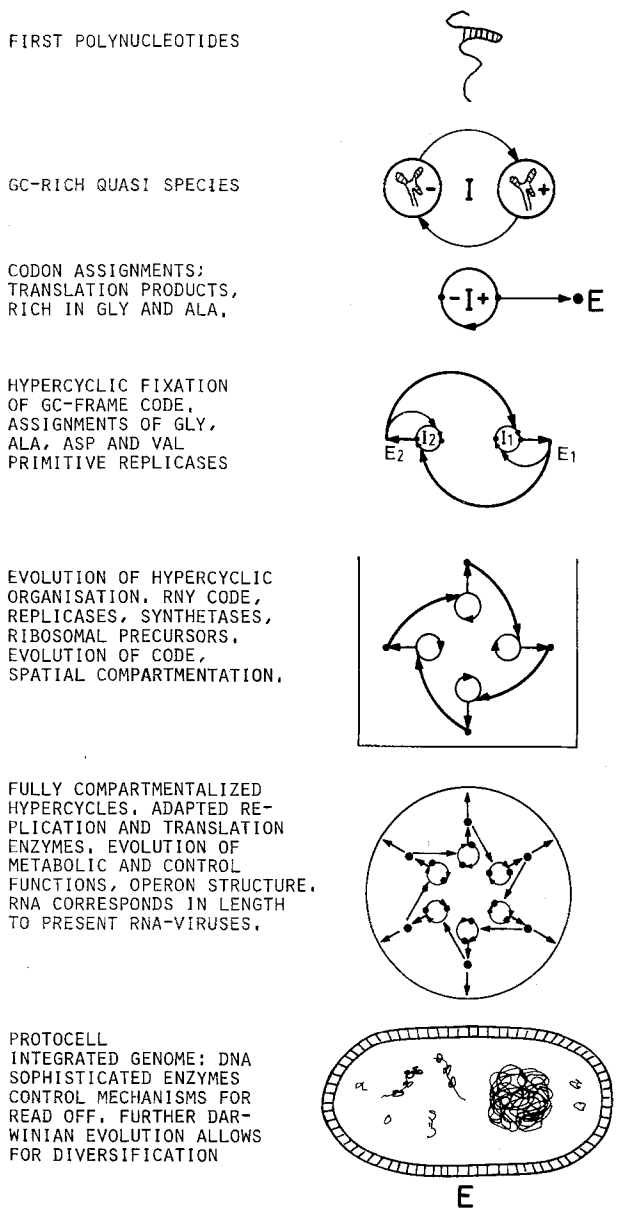


Fig. 63. Hypothetical scheme of evolution from single macromolecules to integrated cell structures

complementary patterns (possibly being the minus strands of messengers) represent suitable adapters. The first amino acids are assigned to adapters according to their availabilities. Translation products look monotonous, since they consist mainly of glycine and alanine residues. The same must be true for the bulk of noninstructed proteins.

4. If any of the possible translation products offers catalytic support for the replication of its own messenger, then this very messenger may become dominant in the quasi-species distribution and, together with

its closely related mutants, will be present in great abundance. The process may be triggered by some of the *noninstructed* environmental proteins, which in their composition reflect the relative abundance of amino acids and hence may mimic primitive *instructed* proteins in their properties.

5. The mutants of the dominant messenger – according to the criteria for hypercyclic evolution – may become integrated into the reproduction cycle, whenever they offer further advantages. Thus hypercyclic organization with several codon assignments can build up. Such a hypercyclic organization is a prerequisite for the coherent evolution of a translation apparatus. More and more mutants become integrated, and the steadily increasing fidelities will allow a prolongation of the sequences. Different enzymic functions (replicases, synthetases, ribosomal factors) may emerge from joint precursors by way of gene duplication and subsequently diverge. Units, including several structural genes, i.e., which are jointly controlled by one coupling factor.

6. The complex hypercyclic organization can only evolve further if it efficiently utilizes favorable phenotypic changes. In order to favor selectively the corresponding genotypes, spatial separation (either by compartmentation or by complex formation) becomes necessary and allows selection among alternative mutant combinations. Remnants of complex formation may be seen in the ribosomes.

We do not know at which stage such a system was able to integrate its information content completely into one giant genome molecule. For this a highly sophisticated enzymic machinery was required, and the role of information storage had to be gradually transferred to DNA (which might have happened at quite early stages).

These glimpses into the historical process of precellular evolution may suffice to show in which direction a development, triggered by hypercyclic integration of self-replicative molecular units, may lead, and how the developing system may finally converge to give an organization as complex as the prokaryotic cell. We want to stress the speculative character of part C. The early phase of self-organization left traces, but no witnesses, so that many important steps still remain in the dark.

It was not even our intention to uncover historical truth. For a process so largely dependent upon chance – where indeterminate microscopic events, such as mutations, amplify and finally determine the course of macroscopic development – a complete reconstruction of history is not possible at all. Even in biology there is a 'poverty of historicism'. On the other hand, the principles governing the historical process of evolution – even in their finer details – may well be susceptible to our understanding. The traces of history in present systems may provide enough clues to allow one day the construction of 'those n equations for the n unknowns'.

All we wanted to show in this part is that the unique class of reaction networks, which we have termed hypercycles, is indeed the simplest realistic molecular organization that ensures the coexistence of functionally related self-replicative units. Self-replication is required for the conservation of information. Hence the hypercycle is the simplest system that can allow the evolution of reproducible functional links. It can originate from one self-replicating unit and its mutants, i.e., from a single (molecular) quasi-species. Its emergence was inevitable, whenever the conditions laid down by Nature allowed it. And yet:

"If anyone can name a more beautiful triangle underlying the composition of bodies, we will greet him not as an opponent but as a friend in the right."
(Plato, Timaios) [97]

This work was greatly stimulated by discussions with Francis Crick, Stanley Miller, and Leslie Orgel; which for us meant some 'selection pressure' to look for more continuity in molecular evolution. Especially helpful were suggestions and comments by Ch. Biebricher, I. Epstein, B. Gutte, D. Pörschke, K., Sigmund, P. Woolley, and R. Wolff.

The work at Vienna was supported by the Austrian 'Fonds zur Förderung der wissenschaftlichen Forschung' (Project Nr. 3502). Ruthild Winkler-Oswatitsch designed most of the illustrations and was always a patient and critical discussant.

Thanks to all for their help.

60. Woese, C.R.: Nature *226*, 817 (1970)
61. Fuller, W., Hodgson, A.: ibid. *215*, 817 (1967)
62. Crick, F.H.C.: J. Mol. Biol. *38*, 367 (1968)
63. Miller, S.L., Orgel, L.E.: The Origins of Life on Earth. Englewood Cliffs, N.J.: Prentice Hall 1973
64. Oró, J., Kimball, A.P.: Biochem. Biophys. Res. Commun. *2*, 407 (1960); Oró, J.: Nature *191*, 1193 (1961)
65. Lewis, J.B., Doty, P.: ibid. *225*, 510 (1970)
66. Uhlenbeck, O.C., Baller, J., Doty, P.: ibid. *225*, 508 (1970)
67. Grosjean, H., Söll, D.G., Crothers, D.M.: J. Mol. Biol. *103*, 499 (1976)
68. Coutts, S.M.: Biochim. Biophys. Acta *232*, 94 (1971)
69. Kvenvolden, K.A., et al.: Nature *228*, 923 (1970)
70. Oró, J., et al.: ibid. *230*, 105 (1971); Cromin, J.R., Moore, C.B.: Science *172*, 1327 (1971)
71. Rossman, M.G., Moras, D., Olsen, K.W.: Nature *250*, 194 (1974)
72. Walker, G.W.R.: BioSystems *9*, 139 (1977)
73. Biebricher, Ch., Eigen, M., Luce, R.: in preparation
74. Erhan, S., Greller, L.D., Rasco, B.: Z. Naturforsch. *32c*, 413 (1977)
75. Dayhoff, M.O.: Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 2, p. 271 (1976)
76. Jukes, T.H.: Nature *246*, 22 (1973); Holmquist, R., Jukes, T.H., Pangburn, S.: J. Mol. Biol. *78*, 91 (1973)
77. Fox, G.E., et al.: Proc. Nat. Acad. Sci. USA *74*, 4537 (1977)

78. Uramer, F.R., et al.: J. Mol. Biol. *89*, 719 (1974)
79. Chothia, C.: ibid *75*, 295 (1973)
80. Chou, P.Y., Fasman, G.D.: ibid. *115*, 135 (1977)
81. Levitt, M., Chothia, C.: Nature *261*, 552 (1976)
82. Gutte, B.: J. Biol. Chem. *252*, 663 (1977)
83. Rigler, R.: personal communication; Rigler, R., Ehrenberg, M., Wintermeyer, W., in: Molecular Biology, Biochemistry and Biophysics Vol. 24, p. 219 (ed. I. Pecht, R. Rigler). Berlin-Heidelberg-New York: Springer 1977
84. Olson, T., et al.: J. Mol. Biol. *102*, 193 (1976)
85. Crothers, D.M., Seno, T., Söll, D.G.: Proc. Nat. Acad. Sci. USA *69*, 3063 (1972)
86. Biebricher, Ch., Druminski, M.: ibid (submitted)
87. Wahba, A.J., et al.: J. Biol. Chem. *249*, 3314 (1974)
88. Blumenthal, T., Lauders, T.A., Weber, K.: Proc. Nat. Acad. Sci. USA *69*, 1313 (1972)
89. Biebricher, Ch.: in preparation
90. Meinhardt, H., in: Synergetics, p. 214 (ed. H. Haken), Berlin-Heidelberg-New York: Springer 1977
91. Fox, S.W., Dose, K.: Molecular Evolution and the Origin of Life. San Francisco: Freeman 1972
92. Dose, K., Rauchfuss, H.: Chemische Evolution und der Ursprung lebender Systeme. Stuttgart: Wissensch. Verlagsges. 1975
93. Onsager, L., in: Quantum Statistical Mechanics in Natural Sciences, p. 1 (ed. B. Kursunoglu). New York-London: Plenum Press 1973
94. Lasaga, A.C., Holland, H.D., Dwyer, M.O.: Science *174*, 53 (1971)
95. Schmidt, E.: Z. Ges. Eis- u. Kälteindustrie *44*, 163 (1937)
96. Kuhn, H., in: Synergetics, p. 200 (ed. H. Haken). Berlin-Heidelberg-New York: Springer 1977
97. Plato: Timaios
98. Schuster, P., Sigmund, K.: in preparation
99. Urbanke, W., Maass, G.: to be published

Received March 28, 1978

