EVOLUTION OF A GENETIC CODE SIMULATED WITH THE COMPUTER

HANS KUHN and CHRISTOPH KUHN

Max-Planck-Institut für biophysikalische Chemie, Abteilung Molekularer Systemaufbau, D 3400 Göttingen-Nikolausberg, Germany

(Received 19 June, 1978)

Abstract. A simple selforganizing model system of molecules is considered and it is demonstrated by a computer simulation, that a genetic code of 16 elements (aminoacids) can gradually be formed by such a system in the course of many generations. By a number of rare chance events, each suppressing other events of equal a priori probability, a single code results out of an immense number of possible codes of the same a priori probability. The result is discussed in relation to the uniqueness of the genetic code in living systems. The computer simulation emphasizes a particular step in a model pathway discussed elsewhere consisting of many assumed physicochemical steps leading to a genetic apparatus.

It is unclear how a genetic code could have been formed and it is not known why each species among the immense number of biological forms has the same code (Calvin 1969, Dose 1975). For approaching an answer it is of interest to study simple theoretical models. Using a simple selforganizing model system we demonstrate the gradual formation of a code of 16 elements ('aminoacids' 1 to 16) by a computer simulation. Each possibility of a code has the same probability to appear in the course of the computer simulation, but one code evolves. It depends on chance events which occur during the process.

The computer simulation is concerned with a particular step in a model pathway discussed in preceding papers (Kuhn 1972, 1976, 1977). This model pathway is obtained by starting with some reasonable initial situation, proceeding step by step, each time trying to trace a simple physico-chemical process. The basic question in modelling each step was: how do we avoid piling up replication errors, how do we find physico-chemical conditions by which assemblies of molecules develop to systems of increasing complexity?

The answer is seen in a specific environment structure being the driving force of the process. By this structure, which is periodic in time and many sited in space, evolution is initiated and driven towards a continuously increasing degree of complexity correlated with a continuous expansion of the populated area.

Important stages on that modelled pathway are the development of replicating strands, the development of assemblies of such strands obtained by aggregation, the development of aggregates which in addition have catalytic properties to produce a second class of macromolecules. These systems develop to assemblies in which this catalytic apparatus gets more and more complex and at this stage the situation emphasized in this paper is

Origins of Life 9 (1978) 137–150.

reached. In a later stage another fundamental change in the organization structure is expected to occur and then the main organization structure of the genetic apparatus is achieved.

1. Basic model

In a theoretical model a simple situation is considered and the consequences are analysed. This method of modelling idealized processes in thought experiments has proved to be fruitful in physical chemistry and should be useful in the search for principal mechanisms in selforganization. The detailed picture that must be given for any model case to analyse its consequence should not be misinterpreted. A hypothetical model pathway of evolution must be an unbroken detailed chain. This chain should be considered as a strongly idealized possibility, focussing on main features of the process, and not as a detailed picture for an historic event.

The model that has been considered (Kuhn 1972, 1976, 1977) is based on the idea that short molecular strands obtained by condensation of monomers are stimulated to undergo multiplication and selection. The stimulus is a very special environmental structure, that is periodic in time, caused by the change between day and night, the tides or the seasons. Monomers attached to appropriate chains polymerize, thus forming complementary chains. By appropriate changes of environmental conditions (e.g. increase of temperature) the double strands fall apart. As soon as the original conditions are reached again, complementary chains are again synthesized. By such a distinct periodic environmental change these particular concerted reactions are stimulated again and again. Daughter strands may be slightly changed by inaccuracies in the polymerization at the matrices. Therefore, in the selection phases many different convoluted forms will be present. They will be different in their ability to protect themselves against decay. The fittest will be accumulated in the course of the process.

The spatial structure of the environment stimulates an evolution to higher and higher complexity. Some times a form may be obtained which is able to survive in a region where the present forms cannot survive. Such a form obtained by a fortuitous error in the synthesis of the complementary chain will usually have a slightly increased complexity: an additional qualification requires a more complex structure. By this effect of hiding more complex forms in less hospitable regions a continuous increase of complexity coupled with a continuous extension of the populated area is reached: the slightly more complex form always accumulates in a region where there is no competition with the other forms.

A prebiotic earth presents the stimuli of evolution in the suggested manner: a manysited environmental structure changing periodically in time.

By such considerations the model system discussed below is obtained in many steps. Important stages on that model pathway are:

(1) the evolution of aggregates of replicating molecules. At this stage assemblies of molecules act as individuals, in the sense of being subject to multiplication, mutation and

selection. The selection advantage of aggregates is assumed to be caused by the fact that only correctly reproduced molecules can interlock appropriately and then survive while molecules with convoluted forms changed by unfavourable replication errors disappear. The high rate of replication errors, which makes a selection of single molecules with specific convoluted forms impossible, and which cannot be avoided at this stage, is made ineffective by this trick. Well organized aggregates can reproduce and evolve.

(2) the evolution of particular aggregates which have the property of acting as a catalyser in the formation of a second class of macromolecules. At this stage two functions are partitioned between two classes of macromolecules (the first class transfers genetic information from one generation to the next, the second class is responsible for the feedback of individual and environment). In all earlier stages one class of macromolecules had both functions. The two classes of macromolecules are functionally coupled and as an entity they represent the individual being subject to multiplication, mutation and selection.

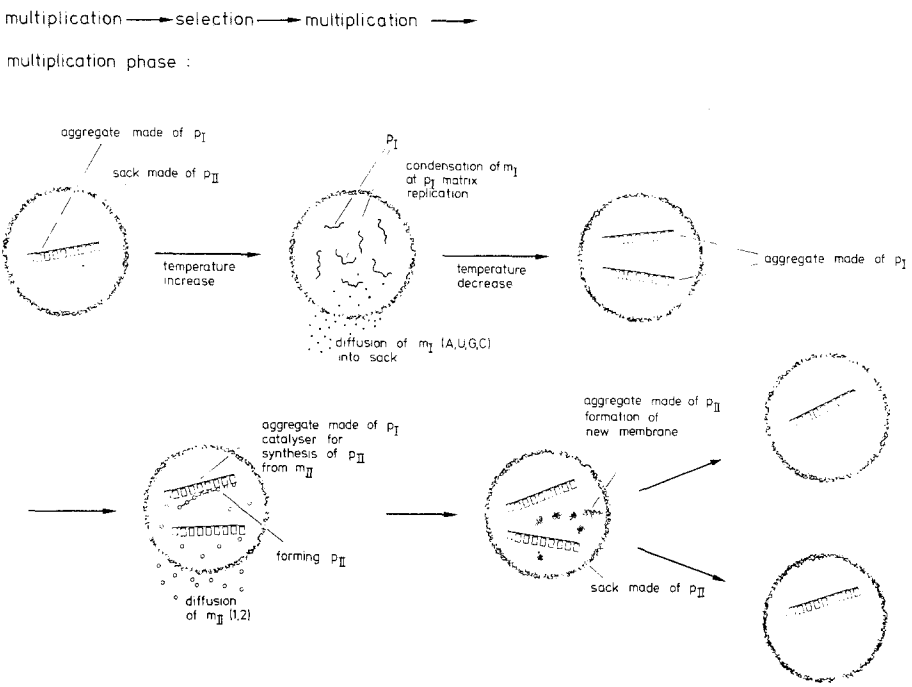

Fig. 1. Model system for computer simulation.

Evolved system consist of aggregates of convoluted strands p_I enclosed in sack of molecules p_{II}. Environmental conditions change periodically in a distinct manner, stimulating particular processes in the system.

Increase of temperature: aggregates of molecules p_I dissociate, monomers m_I diffuse into sack, strands p_I serve as matrix for replication.

Decrease of temperature: aggregates of p_I are formed. Aggregates serve as catalysers for synthesizing polymers p_{II} from monomers m_{II}. Formation of new sacks surrounding p_I aggregates by agglomeration of p_{II} molecules. Selection, continuous repetition of concerted process. In later phases special p_{II} molecules assist in the replication process. A feedback mechanism is initiated by this, which leads to the evolution of the genetic code.

The two classes of interacting polymers are called p_I and p_{II} (Figure 1). The first class p_I (which may be represented by nucleic acids) is the carrier of the genetic information, the second class p_{II} (which may be represented by protein like polymers of aminoacids, known to agglomerate, forming vesicles (Fox 1969)) form a sack. This sack preserves the aggregated macromolecules of the first class from diffusing away.

The formation of such a sack is not unreasonable in the case of nucleic acids and polymers of appropriate aminoacids. Nucleic acids are negatively charged and thus the aggregate is surrounded by a cloud of positively charged counterions. The polymers of aminoacids can have positive and negative charges. At moderate distance this polarizable species will be attracted by the nucleic acid aggregate until it will reach a certain critical distance at which the repulsion from the counterion cloud prevails. The polymers of aminoacids then will accumulate at that distance and agglomerate to form the sack.

As in earlier stages the systems are assumed to be in a periodically changing environment. During the replication phase the molecules of the first class p_I are replicated by attaching monomers m_I (which may be represented by mononucleotides) and bound together. The monomers are present in the surrounding solution. They are small enough to diffuse through the sack of molecules p_{II}, while after polymerization they are imprisoned in the sack. Let us assume that the different chains p_I consist of 4 kinds of monomers as in the biological case and that these monomers are attached to the matrix strand in a complementary way. Following the biological example we call the four kinds of m_I A, U, G, C and assume that A is complementary to U, G complementary to C. In the thought experiment it seems necessary to have two complementary pairs and disadvantageous to have more, thus it appears reasonable why four digits are used in biological systems (Kuhn 1977a).

In the replication phase the chains p_I must be open or weakly hydrogen bonded for allowing the attachment of the monomers, thus the temperature must be relatively high. The next phase is initiated by a decrease in temperature. The chains p_I assume convoluted forms, stabilized by strong hydrogen bonds. The molecules assemble and form aggregates of interlocking molecules. Those p_I molecules that have not succeeded in becoming part of an aggregate are assumed to disappear by diffusing through the sack or by degradation. These aggregates act as catalysers in the polymerization of energy-rich monomers m_{II} (which may be represented by activated aminoacids) to polymers p_{II}. Thus, a synthesis of macromolecules p_{II} takes place, which accumulate and agglomerate in the neighbourhood of each catalyser-aggregate.

In the case considered above of charged aggregates of p_I and polarizable molecules p_{II} forming the sack each additional molecule p_{II} will become part of the sack, the sack will increase in size, become unstable and then stabilize by fission. In this way the original sack is partitioned and new sacks enclosing the catalyser-aggregates are formed from the p_{II} material. Thus two or more daughter individuals occur. The result is a primeval form of a cell division.

In the subsequent selection phase individuals with the more robustious sack have a selectional advantage. They have an increased probability to survive the onset of the next

multiplication phase. In this way a directed evolution takes place. Individuals with the better catalytic apparatus for synthesizing the sack survive and are therefore selected.

Particular p_I aggregates should have a selectional advantage. Such aggregates must quickly build up from single p_I molecules by diffusion and interlocking and they must quickly separate again into p_I molecules. Aggregates with this property may be visualized as consisting of a nucleation molecule and of identical line-up-molecules, the nucleation molecule having a convoluted part and an open chain end (the collector strand), and the line-up-molecules being convoluted similar to present day tRNA's. The convoluted part of the nucleation molecule binds the first line-up-molecule, and the collector strand assists binding others. A catalytic apparatus is assumed to evolve in which each line-up-molecule carries a monomer m_{II} and the monomers bind together forming a molecule p_{II}. The collector strand then acts as primeval messenger, the line-up-molecules as primeval tRNA's.

In the beginning this catalytic apparatus is unprecise, but a strongly directed selection mechanism is present at this stage which leads to the evolution of the catalytic apparatus: individuals where monomers m_{II} interlock with increased precision in the bonding position are able to synthesize macromolecules p_{II} with increased speed. No specificity in the sequence in p_{II} is given and for manufacturing the sack no such specificity is needed (Kuhn 1972).

It can well be imagined that this evolutionary mechanism leads to a system such as the one recently discussed by Crick, Brenner, Klug and Pieczenik (1976). This ingenious speculative model is also based on the idea that the ordering of aminoacids in protein synthesis was accomplished using only messenger-RNA and a few primitive tRNAs, and that the primitive tRNAs were very similar to the present day ones. For realizing a sufficiently strong bonding between the messenger and the particular tRNA molecule to which the growing polypeptide chain is attached it was assumed that the tRNA makes five base pairs with the messenger and that the anticodon loop of each primitive tRNA could take up two configurations. The first configuration is taken up when an aminoacid is attached to the tRNA; it flips to the second configuration when a peptide is attached. The polymerization follows by successively transferring the polypeptide chain to the new aminoacid, and by this flip mechanism the movement along the mRNA of three bases at a time is achieved. Crick et al. concluded that this idea would impose base-sequence restrictions in such a way that A or G should be at the first two positions of each codon, U or C at the third.

Crick et al. find a certain difficulty with their scheme due to the fact that the tRNAs may persist at a wrong position on the messenger, and they mention that the additional binding of the entering tRNA, when in the correct position next to the previous tRNA, would help stabilize this important complex of messenger, tRNA with aminoacid and tRNA with peptide chain.

Crick et al. assumed that these processes started in the primitive soup. The question how this extremely sophisticated machinery could have been reached was not considered. We find it difficult to accept a sudden start in a primitive soup by some improbable

chance event. Thus the aim of our thought experiment is approaching an answer to the question: what can drive molecular systems to start selforganization, to increase complexity and to reach the basic organizational structure of the genetic apparatus?

An essential point in this thought experiment is that a given compartmentation is present (e.g. pores in a rock) which is necessary for hampering diffusion. A system then has a great selection advantage, which is independent of this compartmentation by having a selfmade barrier against diffusion of macromolecules (sack). This idea helps to overcome the difficulty in understanding the evolution of a genetic apparatus requiring enzymes for manufacturing enzymes (hen and egg problem). The difficulty of getting the correct starting condition of the polypeptide synthesis is overcome by assuming a nucleation molecule as part of the messenger initiating the formation of the first peptide bond.

The sack must be permeable for the monomers m_{I} and m_{II} and for those polymers p_{I} which are not incorporated in the aggregate. Therefore a lipid membrane would not be useful at that stage and appears as being a later invention of the evolving systems. Coacervates as part of the environment would not be helpful in that process of evolution of systems, which are increasingly independent by being increasingly complex.

Initial process in the formation of a genetic code

Let us assume that in a certain stage of the evolution of the catalytic apparatus for manufacturing the sack the first two positions on the 'codon' of the collector strand are complementary to the corresponding positions on the 'anticodon' of the line-up-element. Following Crick et al. (1976) we assume that these positions on the codon are restricted to A and G. (We had before considered the formally equivalent case that only one position on the codon is complementary to the corresponding position on the anticodon and that four digits A, U, C, G are possible (Kuhn 1972)). Four kinds of line-up elements then are present which are nonspecific in the monomers m_{II} at that stage. It is assumed that by reaching a certain level of evolution each line-up-molecule has a special cavity to hold its own 'aminoacid'. First these cavities are unspecific and later a distinction between two classes of m_{II} (1 and 2, e.g. hydrophilic and hydrophobic aminoacids) is achieved. Assuming two kinds of cavities, eight kinds of line-up-molecules are present, UU1, UC1, CU1, CC1, UU2, UC2, CU2, CC2, which couple to the collector strand in an arbitrary sequence, restricted by the complementary condition with the corresponding site at the collector strand, e.g.

collector strand	AA ·	GA ·	AA ·	GG ·	AG ·	AG ·	GA ·	GG
line-up-molecules	UU	CU	UU	CC	UC	UC	CU	CC
	2	2	1	2	1	2	1	1

The dots symbolize sites at the collector strand corresponding to the third position in the anticodon triplet.

By some fortunate accident, an individual might possess a collection of line-up-molecules appropriate for coding. However, a genetic code would not be fixed since the

error rate in the replication process would be too high to transfer the necessary information to the next generation. The replication error rate could be diminished if by chance some molecule of class p_{II} would have the property of a replicase (an enzyme which increases the speed of the replication process of molecules p_I and diminishes the replication error rate). Such a p_{II} molecule would have an important additional function.

We ascribe the property of a replicase to a particular sequence of 'aminoacids' 1 and 2. Therefore, if the line-up-molecules are arranged by chance in such a way, that this distinct sequence is obtained, the individual has a selectional advantage. This advantage is usually lost in the next generation. However, a turning point is reached as soon as the following event (or an equivalent event) has happened by chance: an individual occurs which has the line-up-molecules UC1, UU1, CC2, CU2 and not simultaneously any of the other possible kinds UC2, UU2, CC1, CU1 and in which the sequence of the collector strand is such that the replicase is obtained. In this case the code of the two 'aminoacids' 1 and 2 is fixed. The daughter individuals produce again the replicase, if no replication error has occurred. The decreased replication error probability guarantees the multiplication of the replicase-producing form.

Development of the genetic code

At a later stage individuals should be able to produce other simple enzymes besides the replicase. Several aggregates, producing different enzymes, can be imagined to be in the same sack. They would complement each other and with the proceeding evolution process the functioning of the system as an entity would depend more and more on this interplay between parts. To be able to concentrate better on the keyposition of the replicase, we do not consider this fact explicitly in the computer simulation.

At a certain stage the tRNA's are assumed to begin to distinguish between aminoacid 1 and a similar aminoacid 3, or between 2 and 4: the special cavity for 1 (or 2) may be changed sometimes as the result of some replication error and may then form a cavity for 3 (or 4). We will get tRNAs UC3, UU3, CC4, CU4 besides UC1, UU1, CC2, CU2. New replicases will be found by the evolving systems and in this way the genetic code will be gradually formed.

Computer simulation of the development of a code

In the following we assign replicase activity to the sequence 1 1 2 2 1 1 2 2 and assume a decrease of the replication error probability per monomer unit, W, from 1/10 to 1/40 when proceeding to a replicase producing individual. Furthermore, we assume that such an individual has a decay rate, r, reduced to the half:

$$1\ 1\ 2\ 2\ 1\ 1\ 2\ 2 \qquad W = 1/40 \quad r = 1/2 \tag{1}$$

The values are chosen to be reasonable. They can be varied in a considerable range without changing the relevant results of the computation. In each multiplication phase the nucleation molecule with the collector strand is assumed to be duplicated, and three

molecules from each line-up-molecule are assumed to be formed, to have a sufficient pool of line-up-molecules for building up daughter individuals. A chance program decides which line-up-molecule from this pool is introduced at a given position on the collector strand (Figure 2). In case the line-up-molecule chosen first does not fulfill the complementary condition, the computer chooses another line-up-molecule determined by the chance program, and this process is repeated until a complementary line-up-molecule is found. Then the computer proceeds to the next position on the collector strand, and this is repeated up to the last position. Sometimes the computer does not find appropriate line-up-molecules for all positions. Then it stops after 75 trials and the corresponding individual is not counted anymore.

In each period 50 individuals are present at the end of the selection phase. In the multiplication phase we obtain 100 individuals or somewhat less for the reason just indicated. By a chance program it is then decided, which individual is eliminated. That program (developed from earlier programs (Försterling et al. 1971, 1972)) takes care of the fact that the a priori probability of an individual to be eliminated depends on whether it possesses a replicase or not. As mentioned above, the decay probability of each replicase producing individual is assumed to be reduced to $r = 1/2$. This elimination process of individuals is repeated until 50 individuals are left over. Then the whole process of duplication and selection is repeated 100 times.

From then on each tRNA for 1 has a given probability in the reproduction process to convert to a tRNA for aminoacid 3 (probability 0.01) and vice versa, and each tRNA for 2 has the same probability 0.01 to convert to 4, and vice versa. Instead of individuals with replicase (1) we will find after a sufficient number of generations corresponding individuals with 1 or 3 distributed among the positions of 1, and 2 and 4 distributed among the positions of 2, for instance:

collector strand	AA	AA	GA	GG	AG	AG	GA	GA
line-up-molecules	UU	UU	CU	CC	UC	UC	CU	CU
	3	1	4	2	3	1	2	2

The computer is programmed such that these individuals have unchanged properties except individuals with special sequences, which are considered to have a selectional advantage. The following reasonable values are assumed for the decay rate (r) (relative to individuals without replicase) and for the replication error probability per element (W) (1/3 and 2/4 means aminoacid 1 or 3 and 2 or 4 respectively, with the exception of sequence (4)).

1	3	2/4	2/4	1	3	2/4	2/4	$W = \frac{1}{40}$	$r = \frac{1}{4}$	(2)
1/3	1/3	2	4	1/3	1/3	2	'4	$W = \frac{1}{40}$	$r = \frac{1}{4}$	(3)
1	3	2	4	1	3	2	4	$W = \frac{1}{40}$	$r = \frac{1}{8}$	(4)

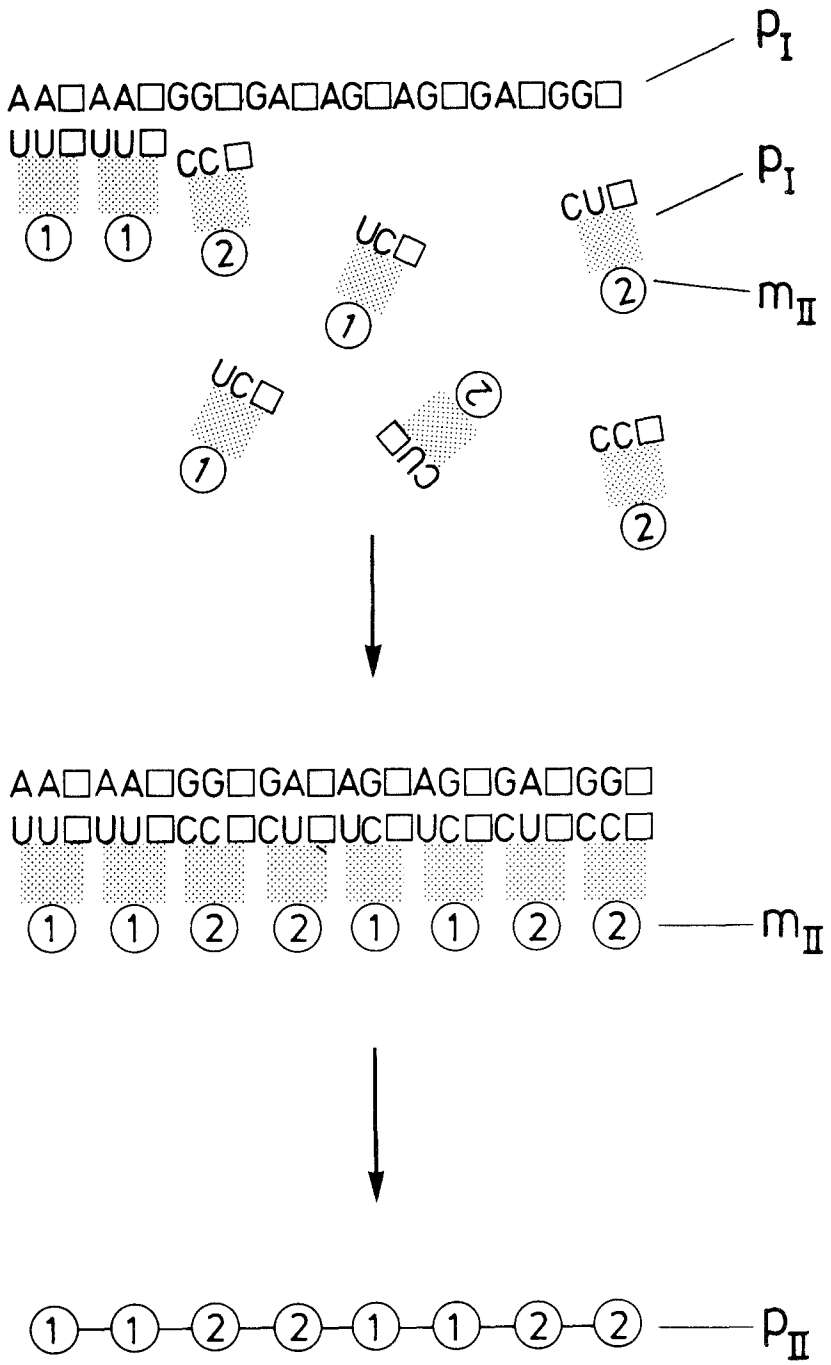

Fig. 2. Scheme of computer simulation.

Formation of aggregate. First two positions in the triplet of line-up-molecule and corresponding positions on collector strand are complementary and restricted to codons A and G. Monomers m_{II} (1 and 2) polymerize to p_{II} with initially statistical sequence of 1 and 2. Individual obtained by chance in generation Nr. 61 (Figure 3).

It has the following properties:

(1) In all line-up-molecules with aminoacid 1 the anticodon is different to all line-up-molecules with aminoacid 2.

(2) The sequence 1 1 2 2 1 1 2 2 is realized, to which we ascribe replicase activity ($W = 1/40$, $r = 1/2$).

As mentioned above, the values $W = 1/40, r = 1/2$ are ascribed to individuals with any other sequence of the form:

1/3 1/3 2/4 2/4 1/3 1/3 2/4 2/4

For taking into account the increase of complexity in the course of the further evolution the computer is programmed to have the following changements in generation number 200: Each strand duplicates into a strand of double length. U and C besides G and A are allowed in the first and second position of the codon and are introduced in the usual manner by accidental errors. New aminoacids 5 and 6, which are similar to 2 and 4 respectively, are introduced in the same manner as 3 and 4 were introduced in generation Nr. 100.

In the computer simulation we assigned diminished decay probabilities r to particular sequences (5) (6) and (7):

1 3 5 4/6 1 3 2 4/6 1 3 2 4/6 1 3 5 4/6 $W = \frac{1}{80}$ $r = \frac{1}{16}$ (5)

1 3 2/5 4 1 3 2/5 6 1 3 2/5 6 1 3 2/5 4 $W = \frac{1}{80}$ $r = \frac{1}{16}$ (6)

1 3 5 4 1 3 2 6 1 3 2 6 1 3 5 4 $W = \frac{1}{80}$ $r = \frac{1}{32}$ (7)

and the previous r value to all other sequences:

1 3 2/5 4/6 1 3 2/5 4/6 1 3 2/5 4/6 1 3 2/5 4/6 $W = \frac{1}{80}$ $r = \frac{1}{8}$

The replication error probability of these forms, W, is assumed to be half as large as in the case of replicase (4), to get practically unchanged reproduction error probability of the individual.

The result of the computation is shown in Figure 3 for an arbitrary example. The first individual with a fixed code has occurred in generation Nr. 61 (Figure 2). It fulfills the two necessary conditions: all elements 1 are coupled with another anticodon than all elements 2 (1 to UC and UU, 2 to CC and CU), and the replicase sequence (1) is present. Thus, aminoacids 1 is coded by AG and AA, aminoacid 2 by GG and GA. In the subsequent generations individuals possessing this code select.

An individual with a code for replicase (2) appeared in generation Nr. 123, selected, and in generation Nr. 169 the first individual coding for replicase (4) appeared, but the form died out again. A new individual coding for replicase (4) appeared in generation 184, succeeded, and the code AA for 1, GA for 2, AG for 3, GG for 4 was then fixed. Another code did not succeed, and this is to be expected, since the number of individuals without code, which could have had the variability to produce sometimes an individual with a new code, was too small for this event to occur with notable probability. Replicase (3) never occurred. The probability that this sequence could have been formed after replicase (2) had selected was too small.

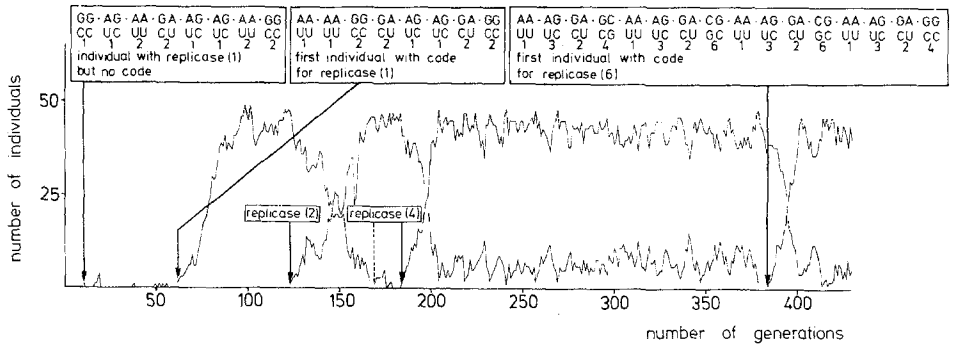

Fig. 3. Result of computer simulation.

Evolution of systems producing replicases (1), (2), (4), (6) in the course of 430 generations. (Replicases (3) and (5) do not appear sufficiently early and have no chance later). The computer is programmed such, that after generation 100 and 200 each time two additional 'aminoacids' are allowed besides the present ones by assuming, that each line-up-molecule changes its aminoacid in the replication phase with a probability of 1%. Up to generation 200 the codon is restricted to A and G and then C and U are allowed besides A and G.

An individual which coded for replicase (6) appeared in generation 384 and forms with code AA, AU for 1, GA for 2/5, AG for 3, GC, GG for 4, CG for 6 were subsequently selected. Replicase (5) (having had no chance after replicase (6) had succeeded) never appeared in the example of Figure 3.

The further computer simulation can be carried out in a similar manner. Aminoacids 7 and 8 which are similar to 1 and 2 respectively, are introduced, later aminoacids 9 and 10, which are similar to 3 and 4 respectively, etc.

In the example shown in Figure 3 for the first 430 generations, the finally resulting individuals have the same code and the same replicase. Each repetition of the computer simulation leads to another code of the same a priori probability.

The computer simulation illustrates that a single code or a few very similar codes result. Each event leading to the codation of an additional aminoacid is a very improbable process and therefore no competitors occur during the selection process of the corresponding form which alone will survive. A later change of the code of a successful form is possible, but is again a rare process: the digit at a given position on the collector strand, and the digit at the corresponding position on the anticodon of the line-up-element must change simultaneously in a complementary manner, and the change must lead to a doublet not yet used for the codation of another 'aminoacid'. In this manner the population slowly drifts to other codes. This can be demonstrated in the computer simulation, if we intentionally omit the addition of new 'aminoacids'.

The fraction of such replication errors, which do not lead to doublets already used for coding, decreases with the number of coded aminoacids. The drift slows down accordingly and a unique code will stabilize when the system will code for 16 'aminoacids' (the maximum number of aminoacids to be distinguished by doublets).

Speculations on the uniqueness of the genetic code

The computer simulation generates a uniform code, and thus the uniqueness of the genetic code seems reasonable for the individuals in a spatial region where the type of evolution considered in the model is proceeding. However, evolution might start at many points on a prebiotic earth and many different codes among all a priori equal possibilities might develop. In the following we intend to show that also in this case a single code should result after a certain time.

As soon as systems with different code (*A*-systems and *B*-systems) meet in the same region, then usually one type will succeed in the competition for the same pool of resources and the other will decay. A host-guest-relation of the two types may develop, but would hold only for a limited time as demonstrated in the following.

The *A*-systems are assumed to enter the region of the *B*-systems. This means at this early stage of evolution considered here that the *A*-systems will have the higher level of organization than the *B*-systems: the degree of organization increases with increasing populated area. The *A*-systems (that have been able to leave their original environment and to reach the area of the *B*-systems) are more sophisticated, being more independent of environment.

The *A*-systems may reach directly the resources in this particular region of the *B*-systems and the *B*-systems will disappear by being displaced. Another possibility is that the *B*-systems, being specialized to their particular region, have the better apparatus than the *A*-systems to make use of special resources, and a host-guest-relation will develop. Thus, the *B*-systems will not be rooted out, but this host-guest-relation cannot be stable: forms of *A*-systems will evolve that possess an enzymatic apparatus which will allow a better adaption to the resources than by the help of the *B*-systems. One enzyme after the other of the *B*-systems will be exchanged for an enzyme, that can be more rationally synthesized by an *A*-system. Host *B* will degenerate and will be finally unnecessary for guest *A* and thus disappear.

A stable coexistence in a host-guest-relation seems only possible at a higher evolutionary level, when the two systems are so complex and different in their structure that they will depend on each other for a long time: the guests have no advantage of being independent of the host and the host-systems are unable to evolve in such a way as to get rid of the guests. This relation determines the later stages of evolution (e.g. plants and animals or animals and viruses). According to the model considered here the struggle for the final genetic code happened before a stage of evolution was reached where a stable coexistence could be possible. The same conclusions can be drawn considering the uniqueness in chirality of living systems (Kuhn 1977a).

Concluding remarks

We have presented a thought experiment for the early evolution leading to a genetic code. The basic model is obtained by assuming

(1) appropriate environmental conditions as initial driving force of evolution causing particular concerted reactions periodically repeated,

(2) an initially given compartmentation enabling the assemblage of molecular systems by interlocking component molecules, and

(3) a many sited environment providing hiding places for forms which are more complex and more clever than the other existing forms.

An evolutionary pathway can be discussed leading to systems with continuously increasing complexity by getting more and more independent of the initial environmental restrictions. Of particular importance is the transition from single molecules to assemblies of molecules acting as individuals in the sense of being subject to multiplication, mutation and selection. By this mechanism of integrating different replicative units by functional linkage a difficulty in understanding prebiotic evolution is avoided, caused by the high rate of replication errors. (A different integration process has been proposed by Eigen (1971, 1976, 1977, 1978) (formation of hypercycles) and the two views were compared (Kuhn 1977b, 1978).)

An important step in the thought experiment is the evolution of systems with a selfmade envelope, which are then independent of environmental compartmentation. A first replicase appears as a by-product in the evolution of a more and more refined apparatus for making this envelope, and this event initiates the evolution of the genetic code.

The aim of such a thought experiment is a simple description of a complex process, not the reconstruction of an historic pathway. We search for the organizational structure of a chemical evolution process and not for detailed chemical aspects. The basic problem is seen in finding appropriate environmental structures in space and time leading to the evolution of functionally ordered systems. In this view the thermodynamic conditions for building up a functional order are obvious and not a basic problem as frequently assumed. The specific model pathway must be speculative. We aim for an unbroken chain of physico-chemical steps and for simplicity of each step. As a model step appears more evident, less mathematization is required to accept it as reasonable.

Such models should be guide lines for experimental work intending the realization of processes which may have been important in early evolution or which may initiate the development of selforganizing chemical systems.

The excellent computer facilities and assistance of the Gesellschaft für wissenschaftliche Datenverarbeitung mbH is acknowledged. We are grateful to Dr. Rolf Landauer, IBM Research Laboratories, Yorktown Heights, for reading the manuscript and for discussions.

References

Calvin, M.: 1969, *Chemical Evolution*, Clarendon Press, Oxford.

Crick, F. H., Brenner, S., Klug, A., and Pieczenik, G.: 1976, *Origins of Life* 7, 389.

Dose, K. and Rauchfuss, H.: 1975, *Chemische Evolution und Ursprung lebender Systeme*, Wiss. Verlagsges. Stuttgart.

Eigen, M.: 1971, *Naturwissenschaften* 58, 465.

Eigen, M.: 1976, *Ber. Bunsengesellschaft* **80**, 1059.
Eigen, M. and Schuster, P.: 1977, *Naturwissenschaften* **64**, 541.
Eigen, M. and Schuster, P.: 1978, *Naturwissenschaften* **65**, 7.
Försterling, H. D. and Kuhn, H.: 1971, *Physikalische Chemie in Experimenten – Ein Prakitikum*, Verlag Chemie, Weinheim, p. 250.
Försterling, H. D., Kuhn, H. and Tews, K. H.: 1972, *Angew. Chem.*, [Internat. Ed.] **11**, 821.
Fox, S. W.: 1969, *Naturwissenschaften* **56**, 1.
Kuhn, H.: 1972, *Angew. Chem.*, [Internat. Ed.] **11**, 798.
Kuhn, H.: 1976, *Naturwissenschaften* **63**, 68.
Kuhn, H.: 1977, in *Biophysik – Ein Lehrbuch*, W. Hoppe, W. Lohmann, H. Markl, H. Ziegler, (ed.) Springer, Berlin, Heidelberg, p. 187.
Kuhn, H.: 1977, in *Synergetics – A Workshop*, H. Haken, (ed.) Springer, Berlin, Heidelberg, New York, p. 69.
Kuhn, H.: 1978, *Physikalische Blätter* **34**, 208, 255.