Computer-modeling origin of a simple genetic apparatus

Christoph Kuhn*

Gruppe Biophysik, Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Hönggerberg, CH-8093 Zürich, Switzerland

Edited by Philip W. Anderson, Princeton University, Princeton, NJ, and approved April 30, 2001 (received for review August 10, 2000)

This computer simulation is based on a model of the origin of life proposed by H. Kuhn and J. Waser, where the evolution of short molecular strands is assumed to take place in a distinct spatiotemporal structured environment. In their model, the prebiotic situation is strongly simplified to grasp essential features of the evolution of the genetic apparatus without attempts to trace the historic path. With the tool of computer implementation confining to principle aspects and focused on critical features of the model, a deeper understanding of the model's premises is achieved. Each generation consists of three steps: (*i*) construction of devices (entities exposed to selection) presently available; (*ii*) selection; and (*iii*) multiplication of the isolated strands (R oligomers) by complementary copying with occasional variation by copying mismatch. In the beginning, the devices are single strands with random sequences; later, increasingly complex aggregates of strands form devices such as a hairpin-assembler device which develop in favorable cases. A monomers interlink by binding to the hairpin-assembler device, and a translation machinery, called the hairpin-assembler-enzyme device, emerges, which translates the sequence of R_1 and R_2 monomers in the assembler strand to the sequence of A_1 and A_2 monomers in the A oligomer, working as an enzyme.

Modern theories of the origin of life are related to M. Eigen's idea that life could have emerged, in principle, by self-organization of matter in a homogeneous phase in a stationary state by intrinsic necessity. The process is bound to the condition of Prigogine and Glansdorff (1–4). These ideas have strongly influenced developments in artificial life and complex systems formation (5). Eigen considers his general theory of molecular Darwinian processes to close the gap between physics and biology.

In contrast, in the view of H. Kuhn (6, 7), understanding the origin of life is not a general but a particular problem: to find a gapless sequence of physicochemical processes, beginning with prebiotically reasonable conditions and leading to systems equipped with a life-like genetic apparatus. Each step must be realistic, detailed, and as simple as possible to emphasize the basic mechanism. The search for such a pathway needs continuous improvement; it is important to have in mind always the entire task and the final goal. The following crucial questions have been addressed (6):

(*i*) What are the boundary conditions that drive a Darwinian process toward increasing complexity, beginning with short molecular strands that (*a*) are reproduced by replication (with occasional variation by copying mismatch) and (*b*) later coagulate into aggregates that are then exposed to selection? How to allow the alternation between these two processes? This task requires an alternation between radically different environmental conditions.

(*ii*) How to avoid the escape of functionally cooperating components and protect against competitors? This aim requires microcompartments that are slightly transparent to allow for occasional escape and results in an occupation of the whole favorable microstructure by the replicating species.

(*iii*) How to drive an evolution to higher complexity? This aim is reached by a diversified environment: populating a new region again and again with slightly less favorable microstructure requires increasing sophistication (bound to increasing genetic information, decreasing replication-error-rate, increasing intricacy of the genetic machinery).

The model pathway of a continuous sequence of steps determined by logical and chemical restrictions is quite narrow: the "blind" trial-and-error process leads by chance, sooner or later, to distinct functional properties allowing a distinct crucial step (e.g., populating a new region where an individual of the new form finds a situation allowing, all of a sudden, fast multiplication: the new form has strong selectional advantage, and further evolution is not suppressed by the old forms, which are now exploitative parasites).

In this view, the emergence of life and self-organization in a homogeneous medium by intrinsic necessity are two totally different processes. Appropriate temporal (cycle of day and night) and spatial (compartmentation as furnished by microporous rocks) environmental conditions (at some very particular location on the prebiotic earth) are of fundamental importance to the emergence of life.

The model (6) was later refined and extended (8–12). Important features of the model are the emergence of hairpins that fold back onto themselves in an antiparallel direction by complementary binding and the subsequent development of aggregates consisting of hairpins linked with their loop to an assembler. These aggregates later lead to a translation device, the *H*airpin *A*ssembler *E*nzyme device (*HAE* device) (Fig. 1). The compact arrangement of hairpins and their binding to the assembler by triplets are reasonable assumptions, as demonstrated by molecular modeling (9, 10), force-field calculations (13), and experiments (14, 15). Basic assumptions of the model are supported by experiments (16–19) and computer models (20, 21).

The computer simulation presented here is significant in facilitating a deeper analysis of the following critical steps, which are not obvious but were intuitively assumed in the theoretical model: (*i*) Can a reading frame develop? (*ii*) Can a binary code develop? (*iii*) Can the binary code be extended to a quaternary code? (*iv*) Can the translation device avoid exploitation by contaminants from earlier stages that are always produced as an undesirable byproduct? (*v*) Can the risk of moving away from the favorable path into harmful side channels be judged?

The Computer Implementation. The theoretical model of H. Kuhn and J. Waser (11) is strongly simplified in the present computer simulation to focus on the significant features. The continuously changing environment requires many parameters (Table 1). The necessary restriction of options (i.e., initial sequences of strands, number of enzymes and their sequences) and the choice of parameter values (i.e., number of compartments, transparency of their walls, length of strands, fitness parameters of devices) are

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: *HAE*, *H*airpin *A*ssembler *E*nzyme.

*E-mail: c.kuhn@access.unizh.ch.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

 www.pnas.org/cgi/doi/10.1073/pnas.141379398

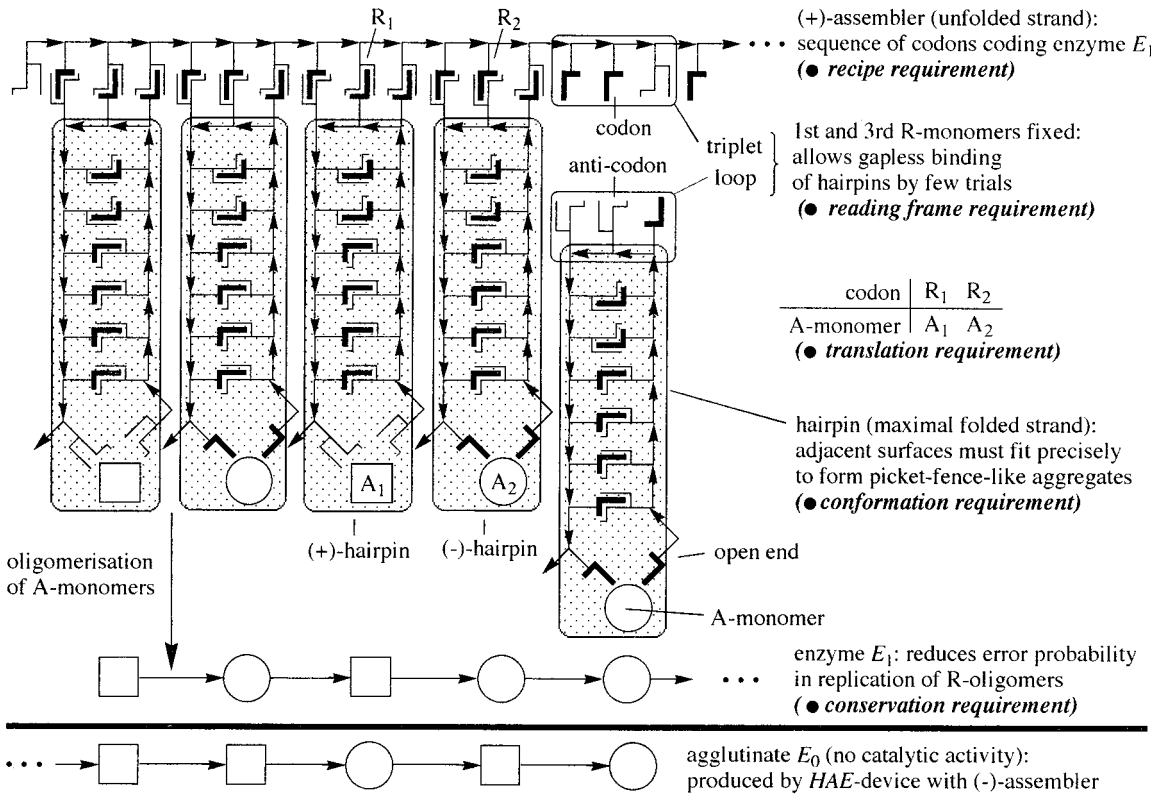

Fig. 1. The *HAE* device. Postulated ancient precursor of a translation apparatus from a nucleotide sequence into an amino acid sequence. On top is the (+)-assembler (only the initial part shown) with five hairpins attached (dotted rounded rectangles; the hairpin on the right is somewhat detached to show its loop), which serve as adapters for A monomers (A reflecting amino acids in biosystems). Hairpins and assembler are strands of R monomers R_1 (white element) and R_2 (black element) binding complementarily (R reflecting ribonucleotide). The direction of the template, (+)-strand and its replica, (−)-strand, are antiparallel (indicated by > and <). The (+)-hairpin and (−)-hairpin differ only in the anticodon (second monomer in loop) and the open end. The (+)-hairpin with anticodon R_2, loop $<R_1R_2R_2<$, and open ends $\{R_1, R_1\}$ is charged with A_1 (square). The (−)-hairpin with anticodon R_1, loop $<R_1R_1R_2<$, and open ends $\{R_2, R_2\}$ is charged with A_2 (circle). The first and third monomer of all adjacent triplets $>R_2R_\alpha R_1>$ along the assembler fixed to R_2 and R_1 constitute a reading frame. The second monomers R_α are codons and by their specific sequence constitute a recipe: the HAE_1 device assists in the synthesis of enzyme E_1, an A oligomer of specific sequence, by translating the recipe on the (+)-assembler with the binary code R_1 for A_1 and R_2 for A_2. Its replicate, the (−)-assembler, leads to an A oligomer E_0 (the terminal part shown at bottom) of unspecific sequence with no catalytic activity, but which helps compartmentation by agglutination. The assembler is the precursor of the messenger RNA, the hairpins the precursors of the transfer RNA's equating R_1 and R_2 with cytidine and guanosine base-paired by triple hydrogen bonds (and equating R_3 and R_4 with adenosine and uridine base-paired by double hydrogen bonds, to appear later). The A oligomers are the precursors of proteins equating A_1 and A_2 with the simplest and most available amino acids glycine and alanine (and equating A_3 and A_4 with aspartate and valine, to appear later).

obviously rather artificial; however, it is not my purpose to simulate a realistic prebiotic situation.

Each successive generation of the cyclic process of a molecular Darwinian evolution consists of three phases (Fig. 2): (*i*) construction of devices presently available simulating formation by diffusion and interlocking; (*ii*) selection; (*iii*) multiplication of isolated strands with occasional variation by copying mismatch. Two R monomers, R_1 and R_2, are covalently linked, forming R oligomers with m monomers. The population of a total of N strands is given as well as the K compartments lined up in a circle. The computer tries to construct devices with strands from the pool of strands in each compartment. The devices obtained are selected by chance. Each device has its specific *a priori* probability S to be selected. The chance of retention increases with increasing sophistication of the device. The selected devices dissociate into approximately N/2 single strands. Each strand has its specific *a priori* probability M to be replicated. If the strand is selected to replicate, random errors violating complementarity occur in the copying process with a probability F per R monomer. At the end of the cycle, the population of N strands has recovered.

In the beginning of the construction phase, each strand is assigned to one of two forms according to its sequence: hairpin (maximal folded strand) or potential assembler (unfolded strand). Then an assembler and a hairpin are chosen randomly from the pool of strands in the compartment singled out. If a triplet at any location along the assembler is found where the hairpin loop is complementary, the hairpin is linked to the assembler, and it acts as a seed for further hairpins to link by complementary binding to the adjacent triplets on the assembler either to the right or to the left, thus without gaps in between. If the hairpin thus fits onto the device, the procedure is continued until the maximum possible number of hairpins link to the assembler eventually forming a *H*airpin *A*ssembler-device (complete *HA* device). But, at the present stage of the construction and at later time, a hairpin that does not fit onto the device is put back into the pool of strands, and another hairpin is chosen randomly from the pool of strands. This is repeated at most ν times (allowing simulation of the repeated transient binding and dissociation in case of missing complementarity). If no fit is achieved within this limited number of trials, the construction is terminated, and an incomplete *HA* device forms. The procedure is repeated for the next assembler, singled out randomly from the pool of strands, and is continued until all strands within the

Table 1. Library of devices C_k of kind k

k	I_k	device C_k	relationship between structure, function and selective advantage	$S(C_k)$	$F(C_k)$	$M(C_k)$	$\nu(C_k)$
Beginning			• R-monomers R_1 and R_2 forming strands which replicate (error probability w = 1/80) by complementary pairing $\{R_1, R_2\}$. • K = 10 compartments, N = 3240 strands, m = 9•3+2=29 monomers per strand. • compartments slightly transparent (stage a): move of strand from compartment n to n ± 1 with probability q = 0.1.				
0	1	open strand	• (+)-strand = template, (-)-strand = daughter strand complementary to template.	1	w	1	6
1	1	isolated hairpin	• hairpin with loop triplet, stem and closed end $\{R_1, R_2\}$ (maximum protection).	2			
2	i+1	incomplete *HA* - device	• assembler with (i = 2, 3, 4, 5, 6, 7, 8) adjacent hairpins attached (additional protection and highly efficient error filters).	3, 4, 6, 8, 12, 16, 24			
3	10	complete *HA* - device	• assembler with 9 adjacent hairpins attached.	32			
→ reading frame.							
Interposition Ip 1			• appearance of two kinds of A-monomers: { - A_1 attach to open end $\{R_1, R_1\}$ and A_2 attach to open end $\{R_2, R_2\}$ of hairpins. - A-monomers do not bind to closed end $\{R_1, R_2\}$.				
4	10	HAE_0 - device	• agglutinate E_0 { - A-oligomer of random sequence or from (-)-assembler of HAE_1. - forms non-transparent envelope (stage b): • if $T_n > T_{max}$ { - division of compartment n, content randomly distributed among pair - elimination of compartment with the least content. T_n = number of devices $C_{k \geq 3}$ in compartment n, T_{max} = 36.	64	w	1	20
5	10	HAE_1 - device	• replicase E_1 { - A-oligomer of specific sequence coded by (+)-assembler - acts additionally on all strands P_i within compartment.	64	w / 3		
6	10	HAE_2 - device	• synthetase E_2 = A-oligomer of specific sequence coded by (+)-assembler.	128	w		
7	2·10	HAE_1 -HAE_2 cooperative	• synthetase E_2 acts on both devices of cooperative. • replicase E_1 assists synthetase E_2.	256	w / 3		100
→ binary code and recipes for enzymes.							
Interposition Ip 2			• appearance of two new kinds of R-monomers: R_3 similar to R_1 and R_4 similar to R_2 with complementary pairing $\{R_3, R_4\}$. • appearance of two new kinds of A-monomers: A_3 attach to open end $\{R_3, R_3\}$ and A_4 attach to open end $\{R_4, R_4\}$ of hairpin.				
8	10	HAE_0 - device	• increasingly improved agglutinate E_0 (additional protection) by incorporating (1,2,3,4) monomers A_3 or A_4.	80, 96, 112, 128	w	1	100
9	2·10	HAE_1 -HAE_2 cooperative	• increasingly replication speed by incorporating (1,2,3,4) monomers R_3 or R_4. • monomer R_3 or R_4 entering reading frame is assumed to be deleterious.	256	w / 3	5/4, 3/2, 7/4, 2	
10	2·10	HAE_1 -HAE_4 cooperative	• synthetase E_4 = improved synthetase E_2.	512	w / 3	2	
→ quaternary code and recipes for enzymes.							

The devices C_k of kind k composed of I_k strands are listed in order of increasing complexity together with assigned values of fitness parameters (the probabilities S of the device to be selected, F of a mismatched replication, M of a strand to be replicated, and the number of trials ν, above which the construction is terminated). Beginning: initial equipment for modeling molecular Darwinian evolution. Interpositions Ip1 and Ip2: accessories appended to initial equipment. Arrows: properties needed to trigger entrance into the stages after the bold line.

compartment are used up. This is done for each compartment. The fitness parameters for each device thus obtained are taken from the library (Table 1).

Setting the Parameter Values. In the library of all possible devices, each device is defined by its functionality, which is given by appropriate values of fitness parameters. Each device is classified into one of the groups indicated by devices C_k of kind k (Table 1).

The Hairpin. A hairpin conformation is highly favored, because the interior is protected against hydrolytic or other attack from the outside. Although copying errors are needed to provide a pool of variations from which selection can take place, errors that prevent hairpin folding are disadvantageous. The length m of R oligomers must be shorter than $1/w$, where w is the bare error probability per monomer, to allow for a sufficient amount of error-free copies, but not too much shorter, to allow sufficient variation ($w < 1/m$, e.g., $m = 29$ and $w = 1/80$).

The *HA* Device and the Reading Frame. An assembler, to which hairpins can attach by complementary bonding, greatly speeds up the formation of interlocking picket-fence-like aggregates of hairpins without gaps in between (Fig. 1). These picket-fence-like aggregates offer additional protection against hydrolysis or other attacks on the sides of hairpins, that are adjacent and thereby provide a strong selective advantage. Furthermore, the *HA* devices are hindered by their size to leave the compartment through pore openings, whereas single hairpins easily disappear by diffusion. This fact justifies assuming a strong selection pressure from hairpin to complete *HA* device (by factor 12; see Table 1; Figs. 3 and 4), but a weak selection pressure (by factor 2.44; see Fig. 5) would be sufficient to allow the evolution of a genetic apparatus. The four possible kinds of triplets (Table 2*a*) along an assembler of random sequence become mutually exclusive for complete *HA* devices: only assemblers with a single kind of adjacent triplets and a population of hairpins with corresponding loops constitute a reading frame, ensuring firm and precise incorporation of hairpins into the complete *HA* device.

The HAE_0 Device and the Agglutinate E_0 Constituting New Boundary of Compartment. The A monomers A_1 and A_2 are attributed with the capability of becoming linked to hairpins: monomers A_1 are attached to hairpins with the open end $\{R_1, R_1\}$, whereas monomers A_2 are attached to hairpins with the open end $\{R_2, R_2\}$. A monomers do not bind to closed-end $\{R_1, R_2\}$. *HA* devices incompletely charged with A monomers are considered

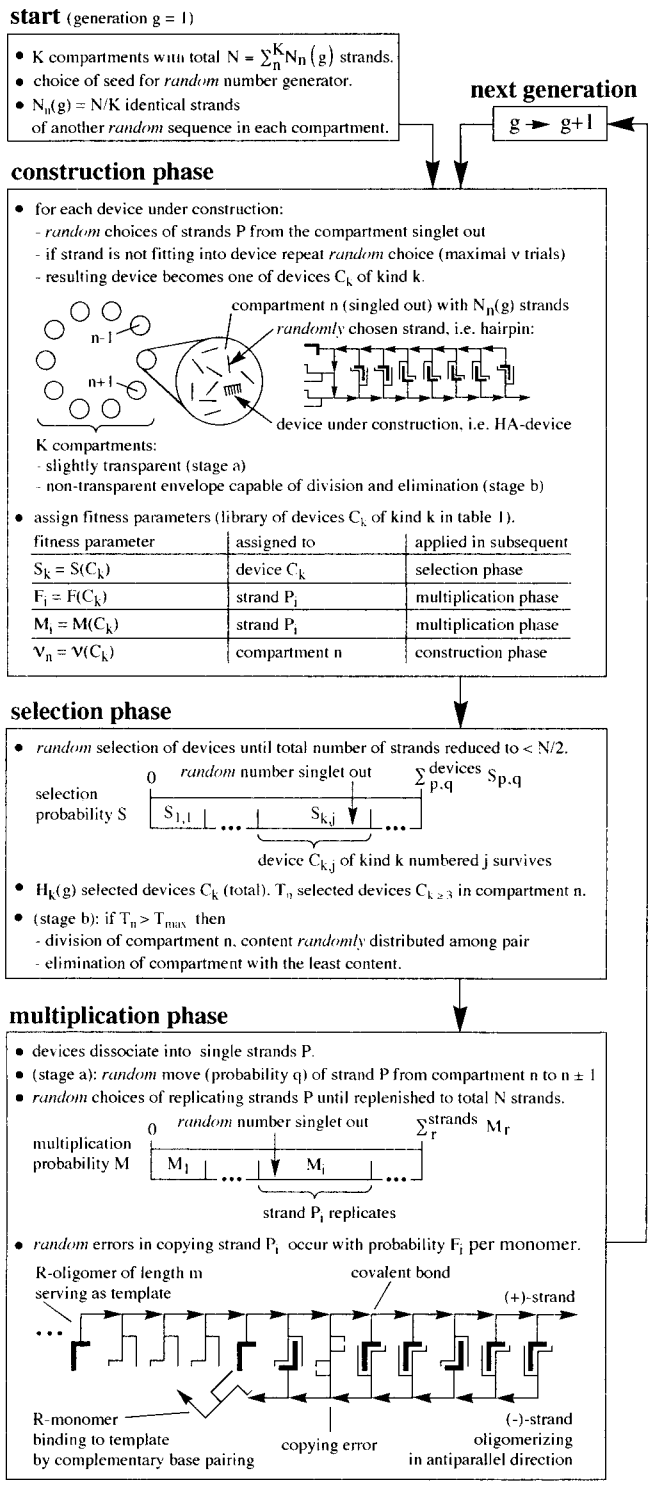

Fig. 2. Implementing the model of molecular Darwinian evolution. Each generation consists of (*i*) construction phase; (*ii*) selection phase; and (*iii*) multiplication phase. The drawings are given as illustration. *Random* numbers are used throughout (specified in italics). Each seed to the random number generator produces a deterministic but statistically uncorrelated sequence of pseudorandom numbers.

to have the same survival chances as *HA* devices. A HAE_0 device is present if each hairpin of the complete *HA* device is charged with an A monomer: the A monomers are then side by side, and their proximity to each other greatly facilitates their becoming

Table 2. List of sequences

a) possible reading frames

triplet	loop
$> R_2 R_\beta R_1 >$	$< R_1 R_\alpha R_2 <$
$> R_1 R_\beta R_2 >$	$< R_2 R_\alpha R_1 <$
$> R_\alpha R_\beta R_\alpha >$	$< R_\beta R_\alpha R_\beta <$
$> R_\beta R_\beta R_\beta >$	$< R_\alpha R_\alpha R_\alpha <$

b) (+)-hairpin and its replica (-)-hairpin:

pair 1: $\left\{ \begin{matrix} < R_1 R_2 R_2 < & \{R_1, R_1\} & A_1 \\ < R_1 R_1 R_2 < & \{R_2, R_2\} & A_2 \end{matrix} \right.$

pair 2: $\left\{ \begin{matrix} < R_1 R_2 R_2 < & \{R_2, R_2\} & A_2 \\ < R_1 R_1 R_2 < & \{R_1, R_1\} & A_1 \end{matrix} \right.$

c) enzyme | list | parent | list | number | sequence

enzyme	list	parent	list	number	sequence
E_1	1 – 6			4	$A_1 A_2 A_1 A_1 A_2 A_1 A_1 A_2 A_2$
E_3	13 – 48	4	31 – 36	33	$A_1 A_4 A_1 A_1 A_2 A_3 A_3 A_2 A_4$
E_2	7 –12			7	$A_2 A_1 A_2 A_1 A_2 A_2 A_1 A_1 A_2$
E_4	49 – 84	7	49 – 54	52	$A_4 A_3 A_2 A_3 A_2 A_2 A_1 A_1 A_4$

d) possible binary codes and their extension to quaternary codes

code 1: →	code 1a:	code 1b:	code 1c:	code 1d:
$R_1 : A_1$	$R_1 : A_1$	$R_1 : A_1$	$R_1 : A_3$	$R_1 : A_3$
$R_2 : A_2$	$R_2 : A_2$	$R_2 : A_4$	$R_2 : A_2$	$R_2 : A_4$
	$R_3 : A_3$	$R_3 : A_3$	$R_3 : A_1$	$R_3 : A_1$
	$R_4 : A_4$	$R_4 : A_2$	$R_4 : A_4$	$R_4 : A_2$
code 2: →	code 2a:	code 2b:	code 2c:	code 2d:
$R_1 : A_2$	$R_1 : A_2$	$R_1 : A_4$	$R_1 : A_2$	$R_1 : A_4$
$R_2 : A_1$	$R_2 : A_1$	$R_2 : A_1$	$R_2 : A_3$	$R_2 : A_3$
	$R_3 : A_4$	$R_3 : A_2$	$R_3 : A_4$	$R_3 : A_2$
	$R_4 : A_3$	$R_4 : A_3$	$R_4 : A_1$	$R_4 : A_1$

(*a*) The four possible reading frames (the two significant reading frames boxed, contrasting the two ambiguous reading frames, which are of no use later on). The triplets with the corresponding loops of hairpins listed ($R_\alpha = R_1$ or R_2 and $R_\beta = R_2$ or R_1 complementary to R_α). (*b*) The four hairpins with anticodon R_1 or R_2 (second monomer in loops $<R_1R_1R_2<$ and $<R_1R_2R_2<$ with given significant reading frame) and A-monomer A_1 or A_2 (attached to open end $\{R_1, R_1\}$ and $\{R_2, R_2\}$) make up the pair 1 and pair 2 of hairpins of mutual replica. (*c*) Among the $2^9 = 512$ possible sequences of A oligomers with length of 9 monomers, A_1 or A_2 are the specific sequences with assumed enzymatic activity of type E_1 (6 sequences, numbered 1–6, sequence 4 shown) and of type E_2 (6 sequences, numbered 7–12, sequence 7 shown). For a given A oligomer, there are $2^9 = 512$ descendants by randomly replacing A_1 with A_3 and A_2 with A_4. The specific sequences are assumed to have enzymatic activity of type E_3 (36 sequences, numbered 13–48, all descendants of E_1, e.g., 6 descendants of parent 4, numbered 31–36, sequence 33 shown) and of type E_4 (36 sequences, numbered 49–84, all descendants of E_2, e.g., 6 descendants of parent 7, numbered 49–54, sequence 52 shown). (*d*) Binary code 1 (or 2) extends to quaternary codes 1a–1d (or 2a–2d), respectively. R_1:A_1 denotes a hairpin that is charged with A_1 and that has the anticodon R_2, which binds complementarily to the assembler with the codon R_1.

linked by covalent bonds (Fig. 1). These A oligomers of random sequence form an agglutinate E_0, replacing the original compartment exemplified by porous rock, thus promoting independence from the environment and providing a strong selective advantage. The compartment is now considered nontransparent but capable of dividing (stage b, Fig. 2 and Table 1), then the devices are arbitrarily distributed over the two daughter compartments, whereas, in keeping the number of compartments *K* constant, the compartment with the least content is assumed to vanish. This procedure simulates the occasional division of the agglutinate into separate entities.

The HAE_1-HAE_2 Cooperative and the Binary Code. An A oligomer of specific sequence with enzymatic activity as of a replicase or synthetase provides a strong selective advantage. There is no premium on the development of a relationship between the R monomer in the middle of a hairpin loop and the A monomer attached to its open end: all four combinations (making up pairs 1 and 2 of hairpins; Table 2*b*) are equally present in HAE_0 devices. But the two pairs are mutually exclusive in a HAE_i device with a recipe for enzyme E_i (a specific sequence of second monomers in consecutive triplets along the assembler): only hairpins of one pair that translate this recipe constitute a code. Let us assume that, by chance, only hairpins of pair 1 are present

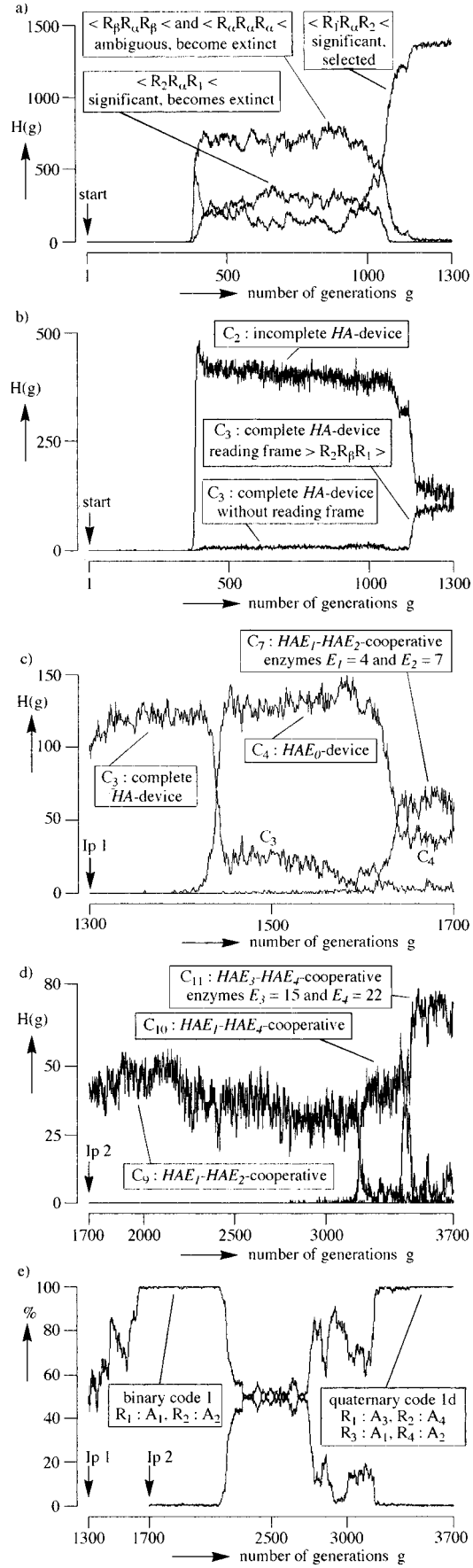

Fig. 3.

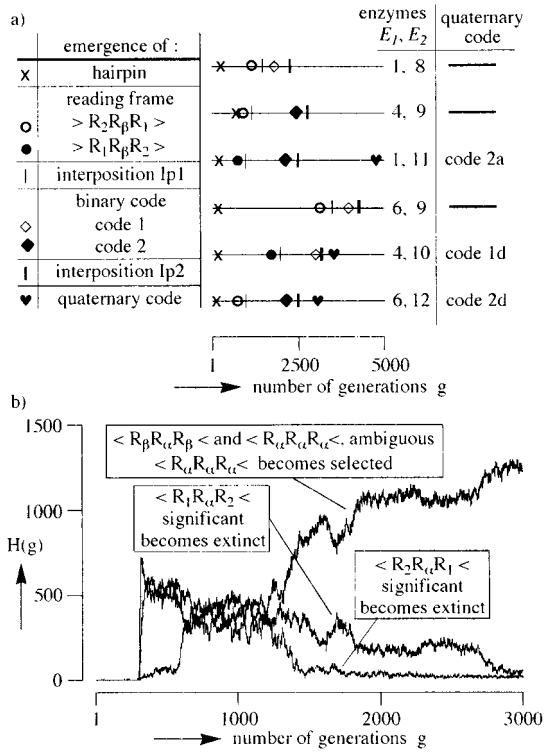

Fig. 4. Simulations with different seeds. Of 20 runs performed, 6 resulted in the emergence of a significant reading frame (50% expected in the average, Table 2*a*). (*a*) In all of these favorable six runs, a HAE_1–HAE_2 cooperative emerged with enzymes E_1 and E_2 from the list 1–6 and 7–12 and with binary code 1 or 2. Only in a fraction (three runs of the six) did the quaternary codes 1a–1d or 2a–2d become manifest within the given number of $g = 5{,}000$ generations. (*b*) Course of computer simulation starting with a random seed that resulted in an ambiguous reading frame contrasting Fig. 3*a*.

in a given compartment and, again by chance, an assembler has the appropriate sequence that allows the production of an A oligomer of a specific sequence of A_1 and A_2 (Table 2*c*; one of

Fig. 3. Computer simulation with a particular seed given to the random number generator. Number H_k (g) of devices C_k at the end of the selection phase versus number of generations g. (*a*) Total number of hairpins shown. At $g \approx 350$, hairpins with all possible loops (Table 2*a*) emerged. At $g \approx 1{,}100$, hairpins with loop $<R_1R_\alpha R_2<$ (belonging to a significant reading frame) dominated. (*b*) Number of *HA* devices shown. As soon as hairpins developed, they formed incomplete *HA* devices. Occasionally, complete *HA* devices with nine adjacent hairpins attached but without occurrence of a reading frame. At $g \approx 1{,}165$, complete *HA* devices with a significant reading frame $>R_2R_\beta R_1>$ were selected (by regions of nucleation). This reading frame exactly was almost the loser shortly before it turned out to be the winner (Fig. 3*a*). (*c*) Interposition Ip1 at $g = 1{,}300$. By chance, the first HAE_0 device appeared at $g \approx 1{,}360$, but only a small fraction of hairpins carried A monomers. At $g \approx 1{,}440$, nearly all hairpins carried A monomers, and the HAE_0 devices took over. At $g \approx 1{,}630$, the HAE_1–HAE_2 cooperative built up with enzymes E_1 and E_1, numbered 4 and 7 (Table 2*c*), simultaneously binary code 1 (Table 2*d*) was selected (by regions of nucleation). (*d*) Interposition Ip2 at $g = 1{,}700$. At $g \approx 3{,}185$, the HAE_1–HAE_4 cooperative producing enzyme E_4 numbered 52 emerged, and quaternary code 1d was selected (Table 2*d*). At $g \approx 3{,}405$, the HAE_3–HAE_4 cooperative producing enzyme E_3 numbered 33 emerged. Enzyme E_3 is a transcriptase (= improved replicase E_1), which assists in the separation of replication and translation by D monomers ("D" reflecting desoxy-ribonucleotide) (6, 11). The nonsensical (−)-assemblers disappear, avoiding accumulation of agglutinate. (*e*) Fraction of coding adapters to total adapters shown. After interposition Ip1, the two pairs of hairpins (Table 2*b*) are of comparable amounts. At $g \approx 1{,}630$, the binary code 1 comes into being and, after interposition Ip2, the quaternary code 1d took over at $g \approx 3{,}185$.

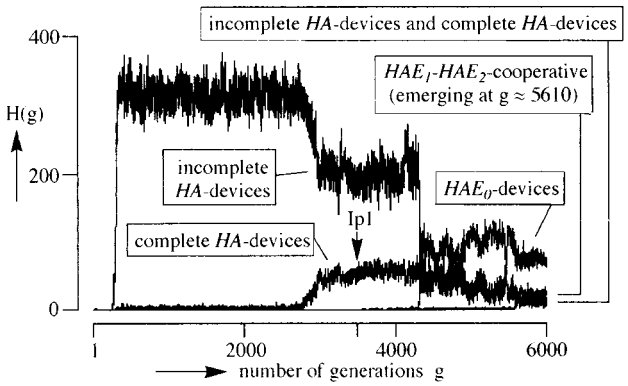

Fig. 5. Simulations with weak selectional pressure. The devices C_k of kind k as listed in Table 1 (k = 0:1:2 (i = 2, 3, 4, 5, 6, 7, 8):3:4:5:6:7) are assigned the following values for probabilities S to be selected (S = 1:1.25:(1.40, 1.56, 1.75, 1.95, 2.19, 2.44, 2.74):3.05:4.58:4.58:5.72:7.15). Incomplete *HA* devices are necessary links in the emergence of complete *HA* devices ($300 < g < 2{,}800$). Incomplete *HA* devices are diminished but still exceed ($2{,}800 < g < 4{,}320$). Contamination is considerable: at g = 3,500, only 34% of the 1,620 given strands are components of complete *HA* devices, 46% of the strands are involved in incomplete *HA* devices, and the residual are 4% free hairpins and 16% open strands. Incomplete *HA* devices fall off and are of no significant contamination as soon as HAE_0 devices are formed and selected ($4{,}320 < g$). The HAE_1–HAE_2 cooperative is stable against exploitation by a great surplus of impurities: at g = 6,000, only about 14% of the given strands are involved in the cooperative.

the possible sequences attributed to act as enzyme E_1). Enzyme E_1 functions as a replication catalyst that reduces the copying error probability to $w/3 = 1/240$. Then the recipe for producing enzyme E_1 is kept over a number of generations but disappears because of the missing selectional advantage of the HAE_1 device. This selection advantage is provided by an A oligomer of another specific sequence of A_1 and A_2 (Table 2*c*; one of the sequences assumed for enzyme E_2). But the synthetase E_2 alone also disappears, because the very copying errors that led to its creation will eventually also destroy it. Then, if in addition to enzyme E_1 an assembler occurs in the considered compartment with a sequence that allows the production of E_2, the HAE_1–HAE_2 cooperative comes into being, which has the power to keep the recipes for E_1 and E_2 and their translation by the pair 1 of hairpins specified by binary code 1 (Table 2*d*).

Extension to the Quaternary Code. R oligomers, with R-monomers R_3 and R_4 incorporated, are assumed to multiply better (by the better-adjusted double- to single-strand transition temperature crucial for maintaining the cycle between the multiplication phase and the selection phase in the given environment). A-monomer A_3 can replace A_1 and A-monomer A_4 can replace A_2 anywhere in the sequence of A oligomers. Then agglutinate E_0 is considered to have an improved function as an envelope. For the enzymes E_1 and E_2, the new A monomers are incorporated without loss of their beneficial properties. Among the pool of possible sequences for equivalents of enzymes E_1 and E_2, it is assumed that for each enzyme there are six descendants with specific sequences forming enzymes E_3 and E_4 each that have the improved enzymatic activity (Table 2*c*). Only a subset of possible quaternary codes is available, e.g., binary code 1 extends to quaternary codes 1a–1d (Table 2*d*).

Results and Conclusion. With a particular seed given to the random number generator, the computer simulation presents devices entering the scene in order of increasing complexity (Fig. 3). With different seeds given to the random number generator for each run, it is shown (Fig. 4) that evolution frequently became stuck at an early stage because of the emergence of an ambiguous reading frame (e.g., Fig. 4*b*) or at the later stage of a binary code because of an incomplete incorporation of the additional monomers R_3 and R_4, A_3 and A_4 within the 5,000 generations given (Fig. 4*a*). This fact supports the idea that many starts of an evolution process were needed until the breakthrough occurred leading to the explosion of life. Fig. 5 demonstrates that even for a small selection pressure, evolution takes place, and the translation device can survive considerable contamination by impurities: proceeding to the next evolutionary stage guarantees protection against exploitation.

The presented computer modeling and the theory on which it is based should be helpful in understanding the evolution of today's genetic code (22, 23). The theory should also be useful in the search for ways to create novel devices by self-organization (24–26).

I am most grateful to the Prof. Peter R. Wills for important suggestions and to Prof. Kurt Wüthrich for his kind hospitality. I also thank Dr. Elizabeth Donley for reading the manuscript.

1. Eigen, M. (1971) *Naturwissenschaften* **58,** 465–523.
2. Eigen, M. (1972) *Nova Acta Leopoldina* **37,** 171–223.
3. Prigogine, I. & Glansdorff, D. (1971) *Thermodynamic Theory of Structure, Stability and Fluctuations.* (Wiley Interscience, New York).
4. Eigen, M. & Schuster, P. (1979) *The Hypercycle—A Principle of Natural Self-Organization* (Springer, Berlin).
5. Kauffman, S. A. (1993) *Origin of Order: Self-Organization and Selection in Evolution* (Oxford Univ. Press, Oxford, U.K.), pp. 357–367.
6. Kuhn, H. (1972) *Angew. Chem. Int. Ed.* **11,** 798–820.
7. Kuhn, H. (2000) *Comp. Biochem.* **41,** 301–362.
8. Kuhn, H. (1976) *Naturwissenschaften* **63,** 68–80.
9. Kuhn, H. & Waser, J. (1981) *Angew. Chem. Int. Ed.* **20,** 500–520.
10. Kuhn, H. & Waser, J. (1983) in *Biophysics*, eds. Hoppe, W., Lohmann, W., Markl, H. & Ziegler, H. (Springer, Berlin), pp. 830–874.
11. Kuhn, H. & Waser, J. (1994) in *Lock-and-Key Principle*, ed. Behr, J.-P. (Wiley, Chichester, U.K.), pp. 247–306.
12. Kuhn, H. & Försterling, H.-D. (2000) *Principles of Physical Chemistry: Understanding Molecules, Molecular Assemblies, Supramolecular Machines* (Wiley, Chichester, U.K.).
13. Kitzing, E. v. (1986) Ph.D. thesis (University of Göttingen, Göttingen, Germany).
14. Lehmann, U. (1985) *Biosystems* **17,** 193–208.
15. Baumann, U., Lehmann, U., Schwellnuss, K., vanBoom, J. H. & Kuhn, H. (1987) *Eur. J. Biochem.* **170,** 267–272.
16. Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **346,** 818–824.
17. Bolli, M., Micura, R. & Eschenmoser, A. (1997) *Chem. Biol.* **4,** 309–320.
18. Luther, A., Brandsch, R. & Kiedrowski, G. v. (1998) *Nature (London)* **396,** 245–248.
19. Joyce, G. F. & Orgel, L. E. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, I. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 49–77.
20. Försterling, H.-D., Kuhn, H. & Tews, K. H. (1972) *Angew. Chem. Int. Ed.* **11,** 821–824.
21. Kuhn, H. & Kuhn, C. (1978) *Origins of Life* **9,** 137–150.
22. Kuhn, H. & Waser, J. (1994) *FEBS Lett.* **352,** 259–264.
23. Ronneberg, T. A., Landweber, L. F. & Freeland, S. J. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 13690–13695. (First Published November 21, 2000; 10.1073/pnas.250403097)
24. Lehn, J. M. (1995) *Supramolecular Chemistry: Concepts and Perspectives* (VCH, Weinheim, Germany).
25. Edelman, G. M. (1987) *Neural Darwinism, the Theory of Neural Group Selection* (Basic Books, New York).
26. Celada, F. & Seiden, P. E. (1992) *Immunol. Today* **13,** 56–65.