Selforganization of Matter and the Evolution of Biological Macromolecules

Manfred Eigen

Die Naturwissenschaften
(The Science of Nature)
1971

Manfred Eigen

- Born: 1927 in Bochum, Germany
- PhD: University of Göttingen, 1951
  - Dissertation on the specific heat of heavy water and aqueous electrolyte solutions
- Max-Planck-Institut für Physikalische Chemie, Göttingen
- Nobel Prize: Chemistry, 1967
  - for his studies of extremely fast chemical reactions, effected by disturbing the equilibrium by means of very short pulses of energy
- His works range from the thermodynamic properties of water and aqueous solutions, and the theory of electrolytes, through thermal conductivity and sound absorption, to fast ionic reactions

I: Introduction

- The question of the origin of life appears as a question of “cause and effect”
- In molecular biology: Which came first, the protein or the nucleic acid?
  - Protein = “function”
  - Nucleic acid = “information”
  - “Function” cannot occur in an organized manner unless “information” is present
  - “Information” only acquires its meaning via the “function” for which it is coding
- It is a closed loop system
  - Although the line from which the loop is formed must have originated somewhere, the starting point will have lost all its importance as soon as the circle is closed
  - Theory of selforganization is required to solve such a problem of interplay between cause and effect
  - Molecular selforganization includes many random events without instructed functional significance
  - What really matters is how certain such random events are able to feed back to their origin and thus themselves become the cause of some amplified action
  - This may build up to a macroscopic functional organization, which includes selfreplication, selection, and evolution to a level of sophistication where the system can escape the prerequisites of its origin and change the environment to its own advantage

The cell is a self-replicating entity

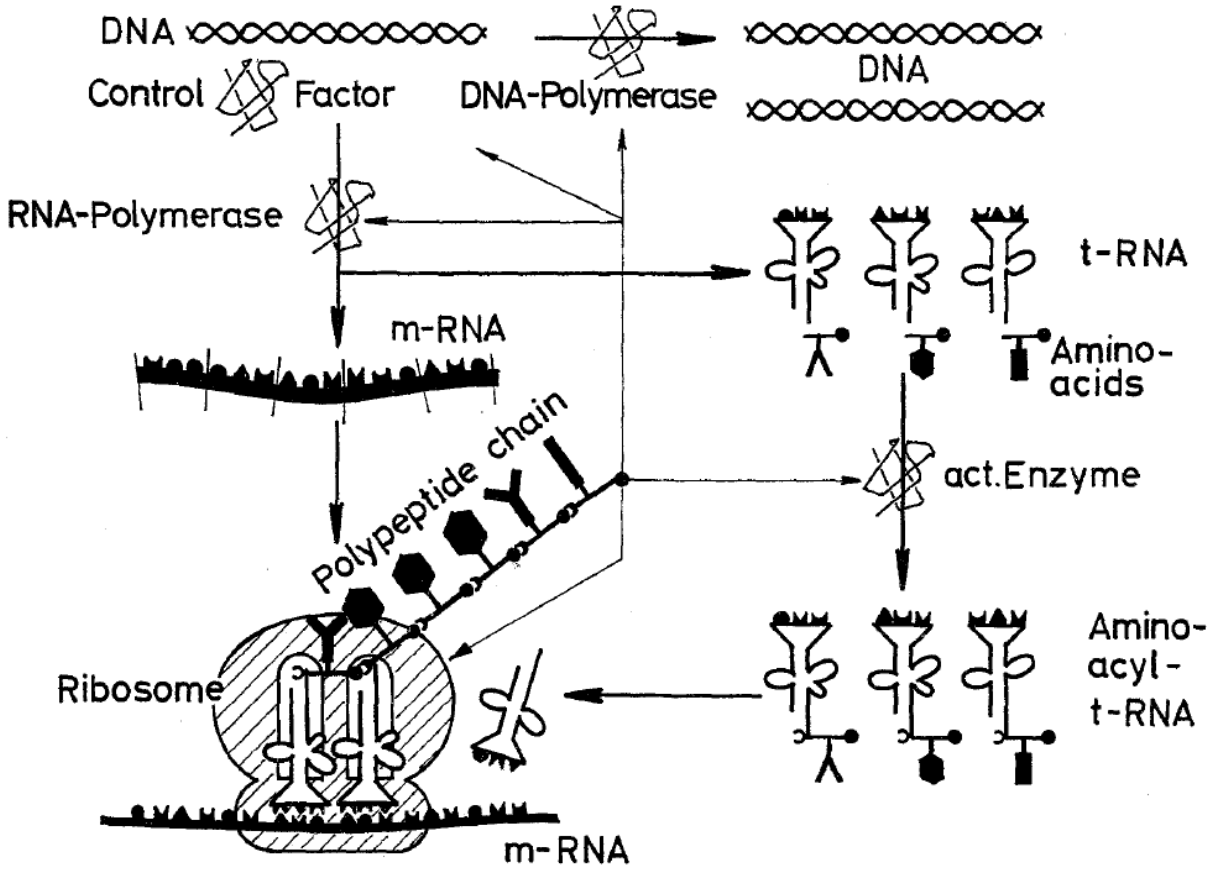


Fig. 1. The selfreproducing biosynthesis cycle of the cell

Nucleic acids and proteins are intimately linked together in their reproduction cycle

How did the origin of life start?

- At the “beginning” there was molecular chaos; no functional organization among the immense variety of chemical species
- Origin of life must have started from random events
- Information theory: instruction requires information
  - Information content:

$$I_1 = K \ln (Z_0/Z_1). \qquad I_{\lambda\nu} = K\nu \ln \lambda .$$

- Selecting a situation with Z_1 out of Z_0 possible outcomes
- γ types of digits to sequences of ν
- Biological macromolecules contain an enormous information capacity
- Nucleotides: γ =4
- Amino Acids: γ =20

Table 2. *The genetic code*

first position	*second position* U	C	A	G	*third position*
U	phe	ser	tyr	cys	U
	phe	ser	tyr	cys	C
	leu	ser	term	term	A
	leu	ser	term	trp	G
C	leu	pro	his	arg	U
	leu	pro	his	arg	C
	leu	pro	gln	arg	A
	leu	pro	gln	arg	G
A	ile	thr	asn	ser	U
	ile	thr	asn	ser	C
	ile	thr	lys	arg	A
	met[a]	thr	lys	arg	G
G	val	ala	asp	gly	U
	val	ala	asp	gly	C
	val	ala	glu	gly	A
	val[a]	ala	glu	gly	G

How big is this information capacity?

Table 3. *ν digits of basis λ have $N_{\lambda,\nu} = \lambda^{\nu}$ possible sequences*

Examples	λ	ν	$N_{\lambda,\nu}$
Small subunits of natural proteins (e.g. M.W. 1200)	20	100	10^{130}
Polypeptides, resulting from AU code only (cf. Table 2)	6	100	10^{78}
DNA chain, coding for 33 amino acids	4	99	10^{60}
AU copolymer coding for 33 amino acids of the AU codon class	2	99	10^{30}
Oligopeptide containing any 12 out of the 20 natural amino acids	20	12	$4 \cdot 10^{15}$
Oligopeptide containing 20 amino acids of the AU codon class	6	20	$4 \cdot 10^{15}$

For comparison: Number of protein molecules of M.W. 10^4

a) assuming closest packing

Universe	10^{103}
1 m thick layer on surface of earth	$2 \cdot 10^{40}$
1 m^3	$6 \cdot 10^{25}$

b) contained in a 10^{-3} M solution (corresponding to a "soup" of appreciable viscosity)

All oceans	10^{42}
pond (100 × 100 m 10 m deep)	$6 \cdot 10^{28}$
puddle (1 liter)	$6 \cdot 10^{20}$

Comparison: age of the Earth = 10^{17} seconds

Conclusion: The information capacity of DNA/RNA/Protein is incredibly large

Introduction of a value parameter

- Information alone is insufficient, there needs to be a “value” parameter to characterize the level of evolution
- “Valuation” characterizes the degree of selforganization of a functional order and defines a gradient for evolution
- Information gains value by selection

- Example:
- Protein of 100AA residues has 10^130 different choices of sequences
- Nature is sophisticated: utilizes parameters such as structural advantages to affect the mechanism of valuation

What properties of matter are required to start self-organization?

Phases of evolution:

1) A prebiotic chemical phase
2) The phase of self-organization to replicating individuals
3) The evolution of individual species

nonliving → living

- "All authors agree on the conclusion that the essential building stones of biological macromolecules – amino acids, energy-rich nucleoside phosphates such as ATP and its base homologs, as well as many other biochemical compounds – could form, where required, and polymerize under prebiotic conditions"

- Catalytic function in combination with various feedback mechanisms causing certain self-enhancing growth properties of the system are one of the decisive prerequisites for self-organization

- Environmental conditions are also required for self-organization, as autocatalytic growth cannot occur in completely or nearly equilibrated systems

- (Equilibrium is a state of maximum entropy)

Thermodynamic aspect of evolution

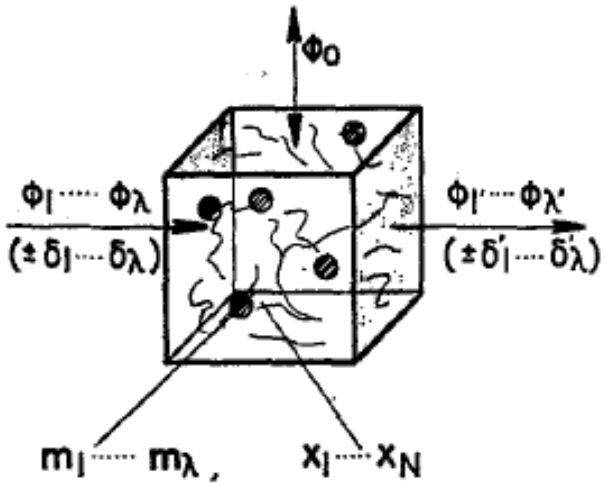
Chemical thermodynamics: the study of the interrelation of heat and work with chemical reactions

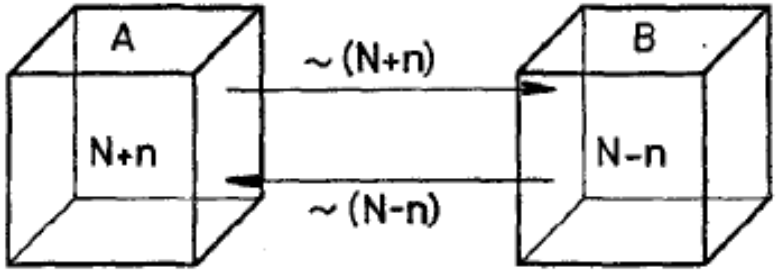
- The energy of the universe is constant.
- In any spontaneous process, there is always an increase in entropy of the universe
- The entropy of a perfect crystal at 0 Kelvin is zero

State functions:

- Internal energy (U): refers to energy contained within the system, while excluding the kinetic energy of motion of the system as a whole and the potential energy of the system as a whole due to external force fields. It keeps account of the gains and losses of energy of the system
- Enthalpy (H): a defined thermodynamic potential, that consists of the internal energy of the system (U) plus the product of pressure (p) and volume (V) of the system
- Entropy (S): is a measure of the number of specific ways in which a thermodynamic system may be arranged, commonly understood as a measure of disorder
- Gibbs free energy (G): a thermodynamic potential that measures the "usefulness" or process-initiating work obtainable from a thermodynamic system

II. Phenomenological Theory of Selection, II.1 the concept of "information"





Depiction of the "information box"

Upper box:

- Energy rich and energy deficient monomeric digits can flow in and out
- Polymeric sequences representing "information" are assembled and decomposed
- Both may be enzyme catalyzed processes
- Steady state is maintained via a control of fluxes/concentrations

Lower box:

- Equilibrium is represented, involving fluctuations +/- n
- In thermodynamic equilibrium there are no net macroscopic flows of matter or of energy, either within a system or between systems
- A system will spontaneously evolve toward thermodynamic equilibrium, the configuration with maximum entropy

Relating the information box to evolution

- It is theorized through a series of phenomenological equations, that each mutation leading to an increase in the selective value corresponds to a negative fluctuation of entropy production, indicating instability of the existing steady state
- Evolution at constant flows corresponds to a sequence of such instabilities, in which the dominant species i_m die out in favor of the new species i_{m+1} according to a finite selective advantage
- $(W^F_{m+1}-W^F_m)$ where W^F is derived as a selective value
- This can be compared to mean productivity (E), where W > E **** (many additional equations and parameters are used to derive this comparison)
- Whenever a mutant with selective advantage (W > E) occurs, it will inevitably outgrow the former distribution
- This gives a physical definition to the Darwinian term "fittest"
    - Darwin states "This preservation of favorable individual differences and variations, and the destruction of those which are injurious, I have called Natural Selection, or survival of the **Fittest**"
    - He notes a physical origin: "It is no valid objection that science as yet throws no light on the far higher problem of the essence or origin of life"

    - <u>Conclusion</u>: Fitness is dependent upon constant fluxes of entropy, leading to instability and selective advantages, leading to evolution

Stochastic (random) approach to selection

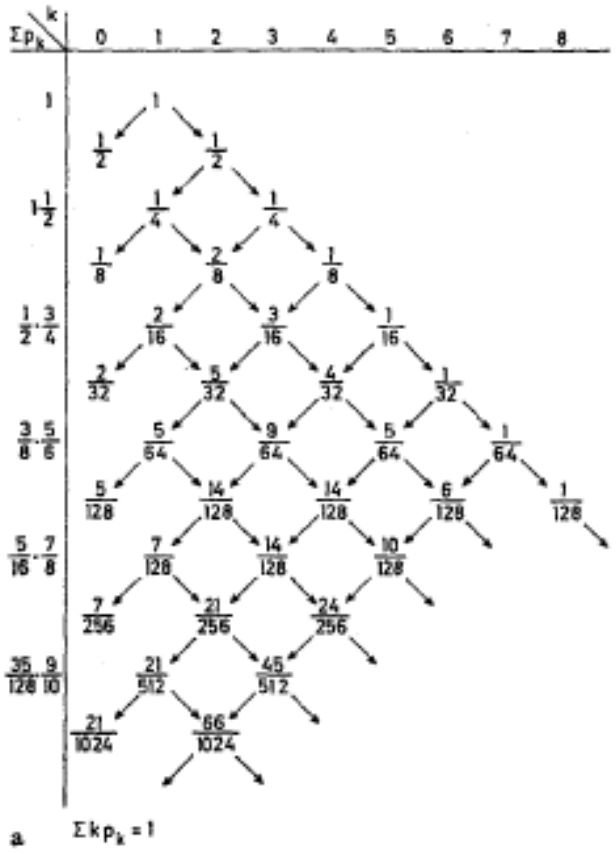
So far, selection has been treated as a deterministic process

Limitations:

1) The elementary process leading to a specific mutant is non-deterministic

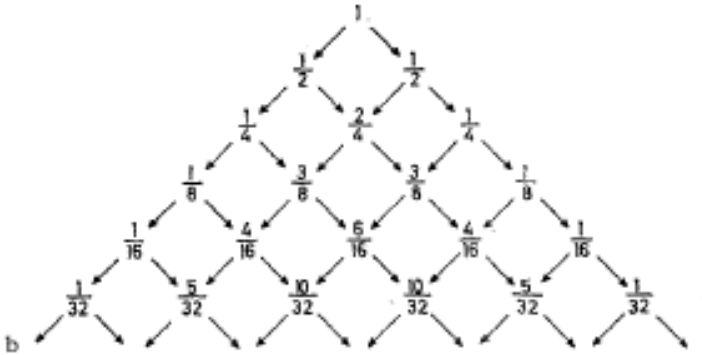
2)The growth process itself is subject to statistical fluctuations

What does selection look like at steady state?



Probabilities in the steady state model (modified Pascal triangle)

- K = redundancy in each copy
- Each step represents a change of the population by one with an equal chance for addition or removal
- Chart represents dissymmetry brought about by extinction



What it means?

- The individual information content narrows down to only a few, or even one, highly redundant information carriers before the total information content is extinguished
- Such selection behavior, caused by the autocatalytic nature of the formation process, represents a typical case of survival of the survivors
- There is no other criterion for selection than the outcome (either survive or don’t)

IV: Self-organization Based on Complementary Recognition: Nucleic Acids

- Complementary instruction is based on exclusive pair formation between A and U or C and G
- Biochemist reasoning: “specific enzymes”
- Physical chemist reasoning: “specific forces”
- Specific complimentary paired structures proposed by Watson and Crick

Why do the bases pair this way?

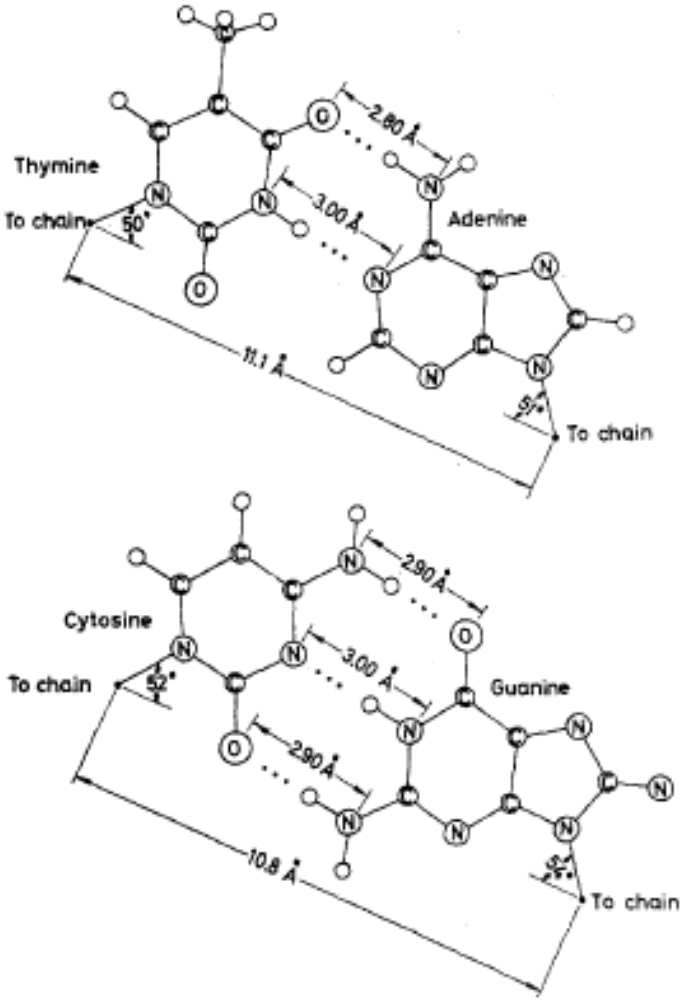


Fig. 8. The complementary base pairs

Other Possibilities

- Hydrogen bonding alone is insufficient
- There is a difference in geometry of the different pairs
- Important so that the isomorphic structure of the two pairs will be advantageous

What are the relative stabilities?

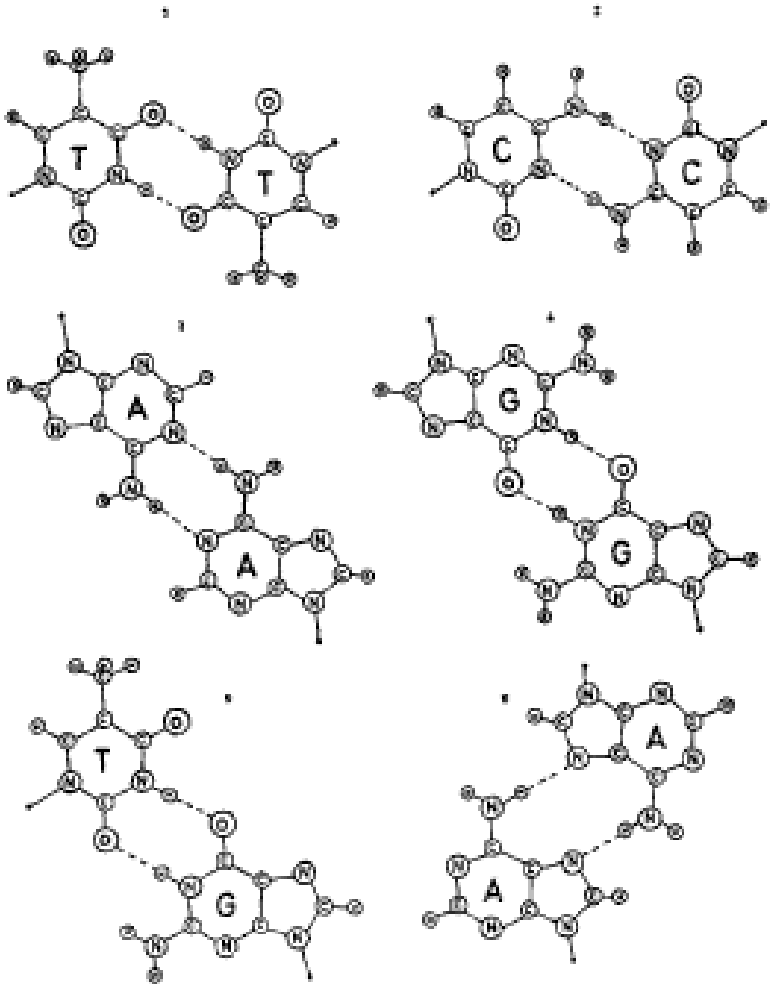


Fig. 9. "Non-complementary" base combinations

Experiments were preformed to assess stability of base pairing combinations

Table 12. *Stability constants of base pairing.* $K_{Ass}[M^{-1}]$ *at 25 °C (2'-3'-5'-O-substituted ribonucleosides in non-polar media)*

CCl_4 / C_6H_6	U	A	C	G
U	45 / 15	550	< 50	$< 10^2$
A	150	22 / 8	< 50	$< 10^2$
C	< 28	< 28	50 / 28	$> 10^4$
G	$< 1.2 \cdot 10^3$	$< 1.2 \cdot 10^3$	$3 \cdot 10^4$	$\sim 10^3$ / $1.2 \cdot 10^3$

- Experiments done by:
  A. Rich, J. Derkosh, T. Funck, R. Hopman, and F. Eggers
- Relative stability measured by adding nucleosides to nonpolar solvents
- Conclusion: The complementary pairs AU and GC are the strongest when compared to the alternatives
- AU is at least ten times more stable than UU or AA
- GC is more stable than CC or GG
- GC is more stable than AU

This preferential complementary base pairing is an important prerequisite of code formation

Codons with three digits are preferred for pairing with anti-codon

Table 13. *Stability constants for pairing of base triplets and quadruplets (tri- and tetranucleotides) with exposed regions (preferably anticodons) of t-RNA according to P. Doty et al. [80, 81]. $K_{Ass}[M^{-1}]$ was measured in aqueous solution of 1.0 M NaCl, 10 mM $MgCl_2$ and 10 mM phosphate at pH 7 and 0 °C. K-values $< 400\ M^{-1}$ are not distinguishable from "no association"*

f-met-*t*-RNA AA[UAC]UC	K_{Ass}	tyr-*t*-RNA AA*[AU*G]UC [a]	K_{Ass}
AUG (regular codon)	1200 ± 200	UAC (regular codon)	700
AUGA	13500	UACA	90000
AUGU	1400	UAU (3′-wobble)	700
AUGC	900	UAUA	37000
AUGG	1000		
GUG (5′-wobble)	1200	phen-*t*-RNA * AY[AAG*]UC*	
GUGA	9800		
GUGU	1000		
		UUC (reg. codon)	900
		UUCA	10000
		UUU (3′-wobble)	300
		UUUA	1000

A* = N(6) dimethyl-A; U* = pseudo-U (ψ); G* = 2-O-methyl-G; C* = 2-O-methyl-C.

[a] Private communication by O. C. Uhlenbeck.

Codons with less than three would be very unstable (at least for A and U)

Codons with more than three become too "sticky" (at least for G and C)

These stability parameters are important because the lifetime of a codon-anticodon pair should not exceed milliseconds so that enzymes with corresponding turnover numbers can adapt optimally

Conclusion: The reason the triplet codon is in place is not only for the coding of at least 20 amino acids

Additional Theorization

Can nucleic acids organize a self-replicating and further evolving unit without catalytic help?

No, such a system would not be able to organize itself into any type of correlated function

What does it need to do so?

V: Self-organization Via Cyclic Catalysis: Proteins

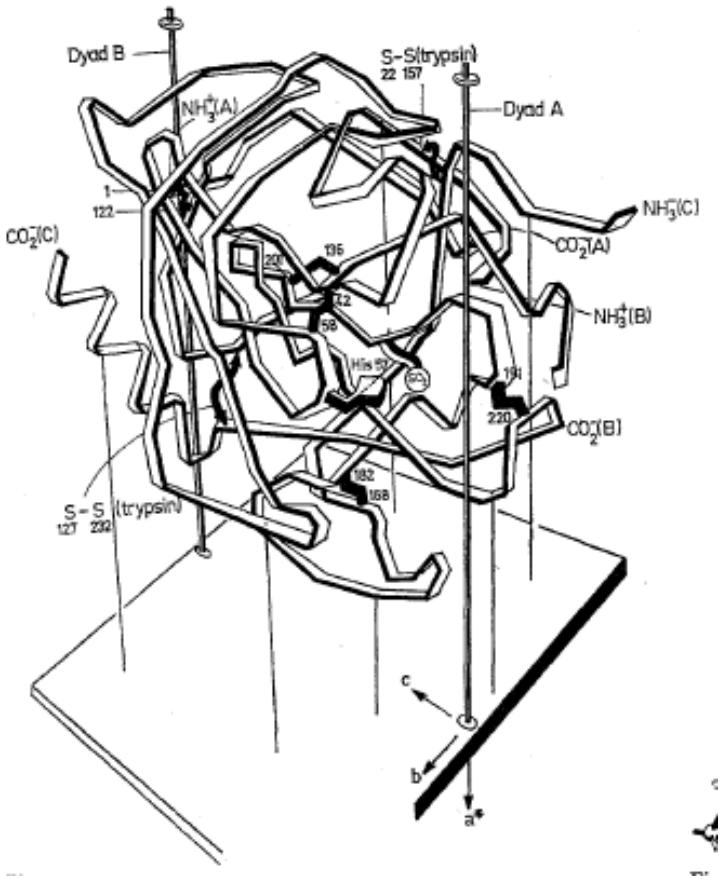
Can proteins alone form self-instructive systems? – No.

- Unlike nucleic acids, there is no property of self- or complementary instruction inherent to amino acids
  - Certain amino acids such as collagen may have a complimentary arrangement of certain amino acids, but it is not an inherent property of the digits

- Proteins lack the very important property of mutagenicity
  - When an error occurs, they would not be able to reproduce their error copy

- Specific spatial folding hinders any straightforward copying process

X-ray structure of chymotrypsin provides excellent example for the dependence of function on a sophisticated structure



- Groups of distant sequential location are brought together into a precisely fixed special arrangement
- Diversity of specific recognition sites is demonstrated by large variety of antibodies
- Active center of chymotrypsin indicates recognition by proteins is a unique result of special folding and not any inherent property of the digits

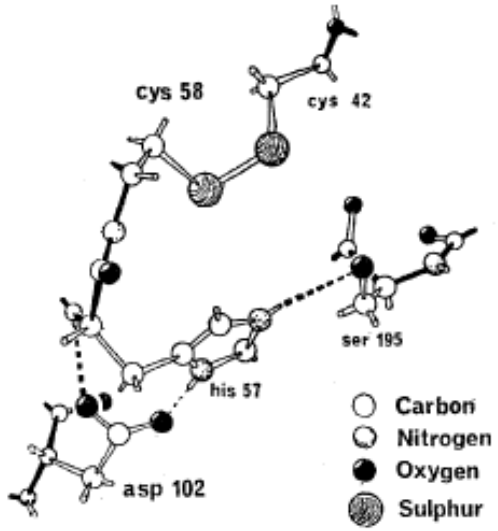


Fig. 10b

Catalytic networks can be made through a series of enzyme catalyzing functions

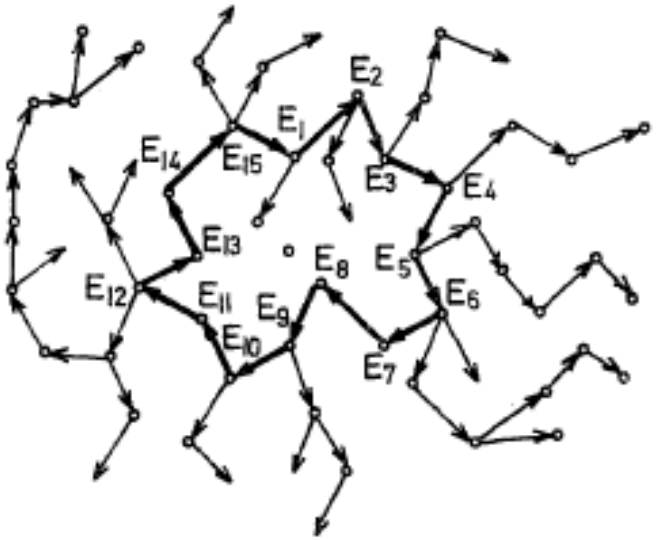


Fig. 11. Catalytic network of proteins, including a closed loop: $E_1, \ldots, E_{15}$

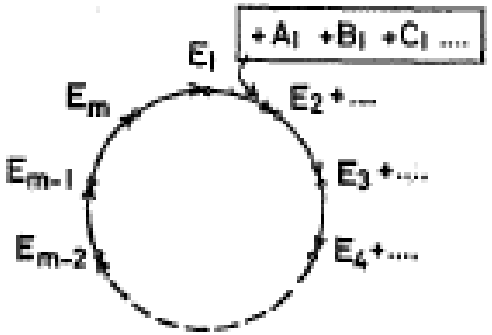


Fig. 12. Graph representation of a catalytic cycle

- A branched catalytic network results from the activation of one catalyst by another

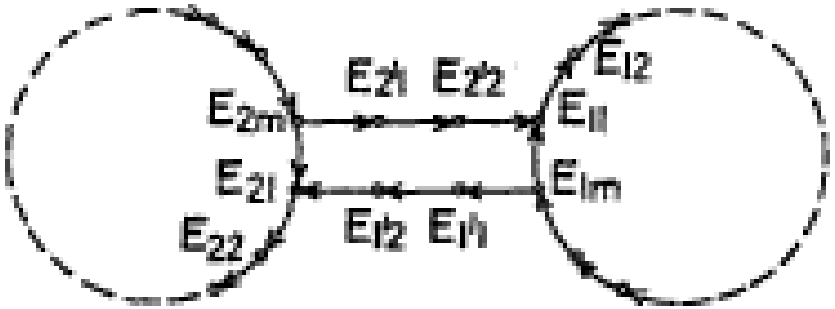
<u>Examples:</u>

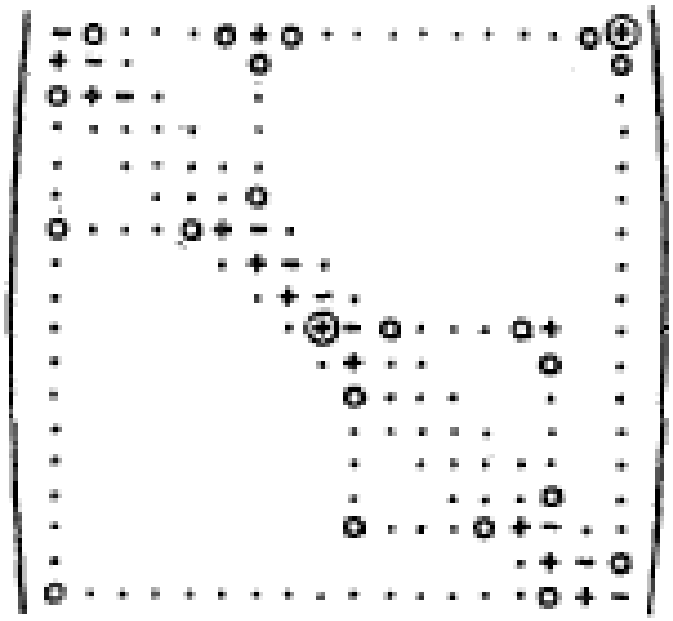
- 1) Activation of chymotrypsin from chymotrypsinogen via enzymatic cleavage of a peptide bond close to one end of the chain
- 2) Certain proteins have the ability to catalyze the condensation of a limited number of amino acids into chains of specified sequences (various pentapeptides)

- Some enzymes have polyfunctional branches to make the network self productive

- If a loop is large enough, all auxiliary functions can easily be located in the branches

A number of different, interacting cycles could exist





Three cycles are shown:

- Two individual cycles and loop involving both cycles
- Due to the feedback, the cycles do not compete for selection, rather stabilizing each other
- The matrix of the rate coefficients of the total system is a reflection of reaction behavior
- Columns=species
- Rows=reactions
- Circled (+) represents coupling points of the loops

Additional Theorization

- Two properties of proteins make them more suitable than nucleic acids for starting self-organization:
  - The higher precision of recognition of certain substrates gained with the help of their tertiary structure
  - The higher information content involved in a multistep cycle (with branches), as opposed to a single chain of restricted length

- The major disadvantage is that proteins which catalyze their own reproduction through specific cycles will not automatically reproduce their mutants resulting from error copying, even if these were to offer advantages

V: Selfordering by encoded catalytic function: The requirement of cooperation between nucleic acids and proteins

- Nucleic acids provide complementary instruction as the basis of selective selfreproduction and code formation using a quaternary digit system
  - Cooperitivity of digit interaction allows for selective and adequately processed codon-anticodon recognition
  - Characteristic single strand structures can be targets for enzymatic recognition

- Proteins provide an enormous functional and recognitive diversity and specificity
  - They can link up via catalytic couplings, increasing information capacity
  - Structure modifications may provide controlling properties

- Nucleic acids provide selective advantages (beneficial mutations) for protein catalytic functions

A Self-reproductive Hyper-Cycle

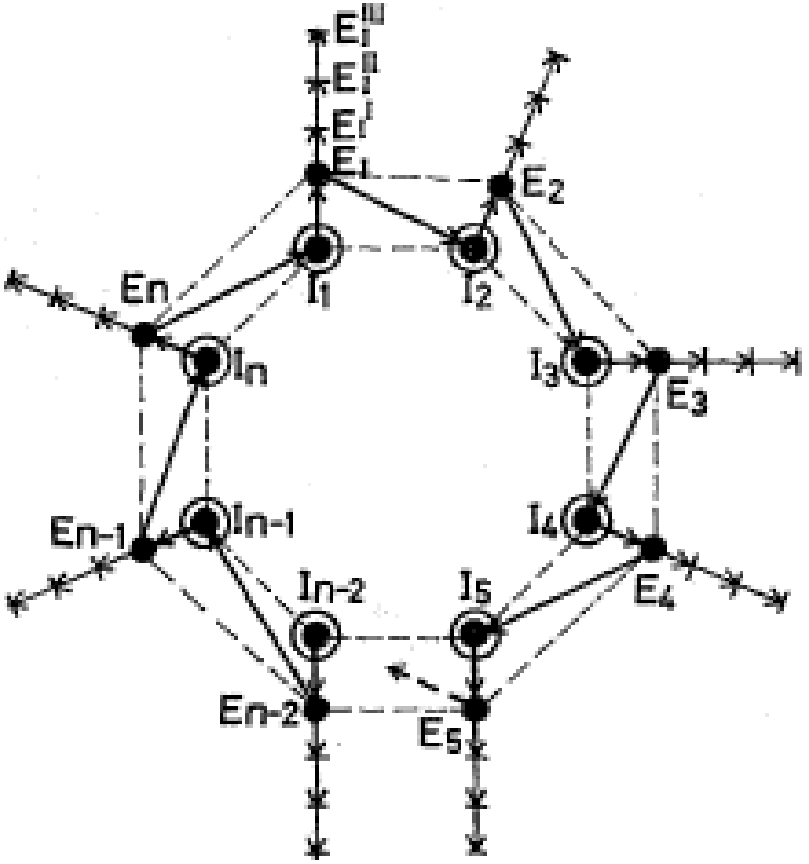


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

1) Each cycle has autocatalytic growth properties

A Self-reproductive Hyper-Cycle

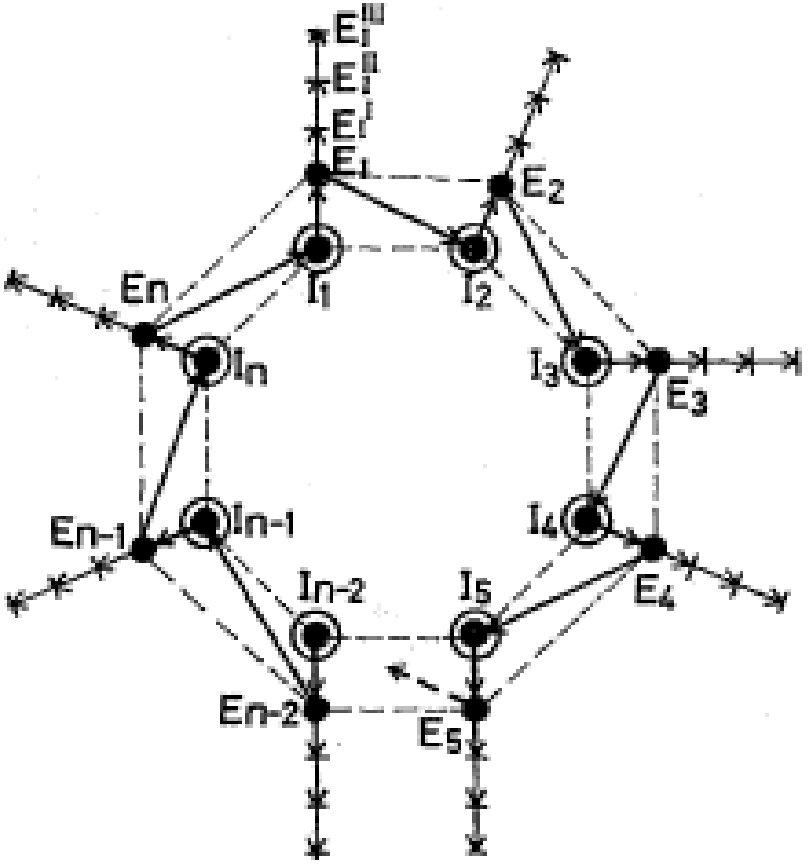


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

2) Independent cycles compete for selection

A Self-reproductive Hyper-Cycle

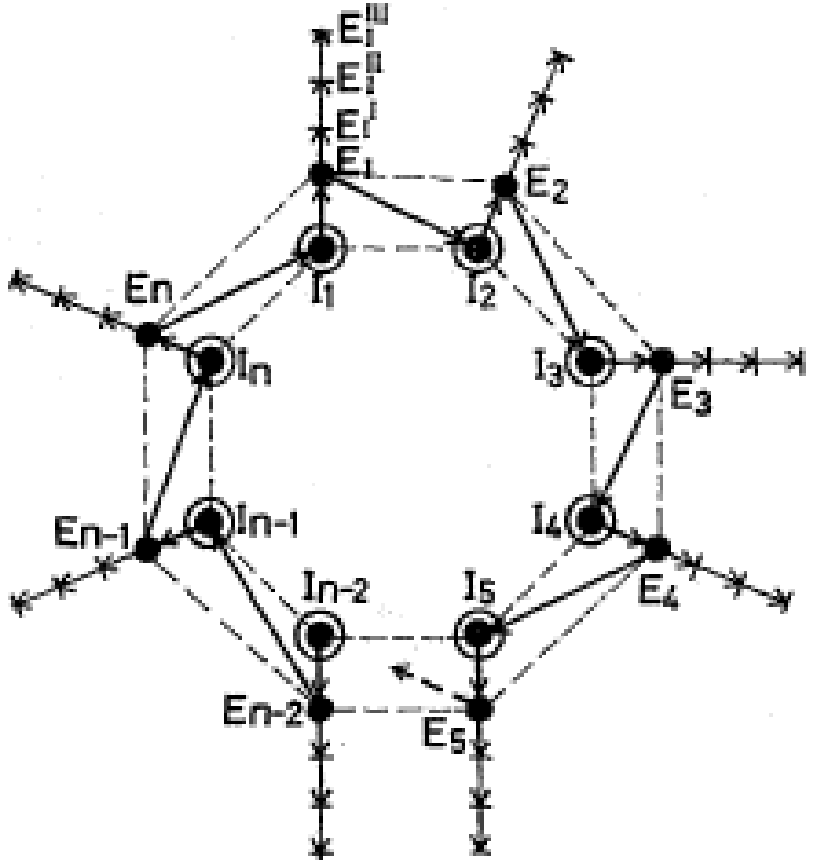


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

3) Selection will be very sharp, possibly resembling "all or none" behavior

A Self-reproductive Hyper-Cycle

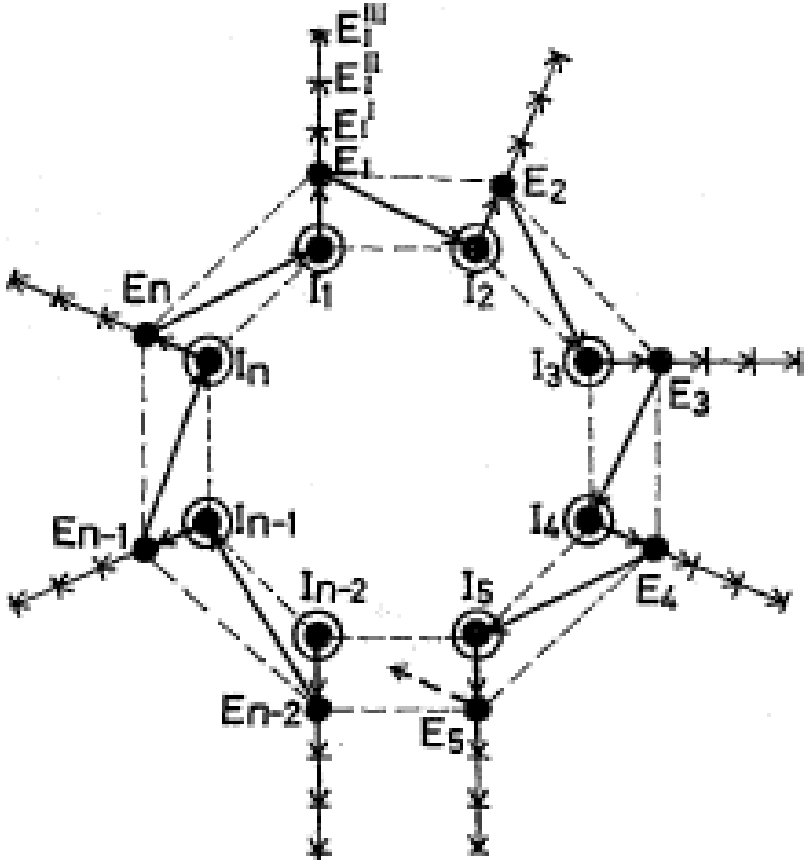


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

4) The system will be able to utilize very small selective advantages and evolve quickly

A Self-reproductive Hyper-Cycle

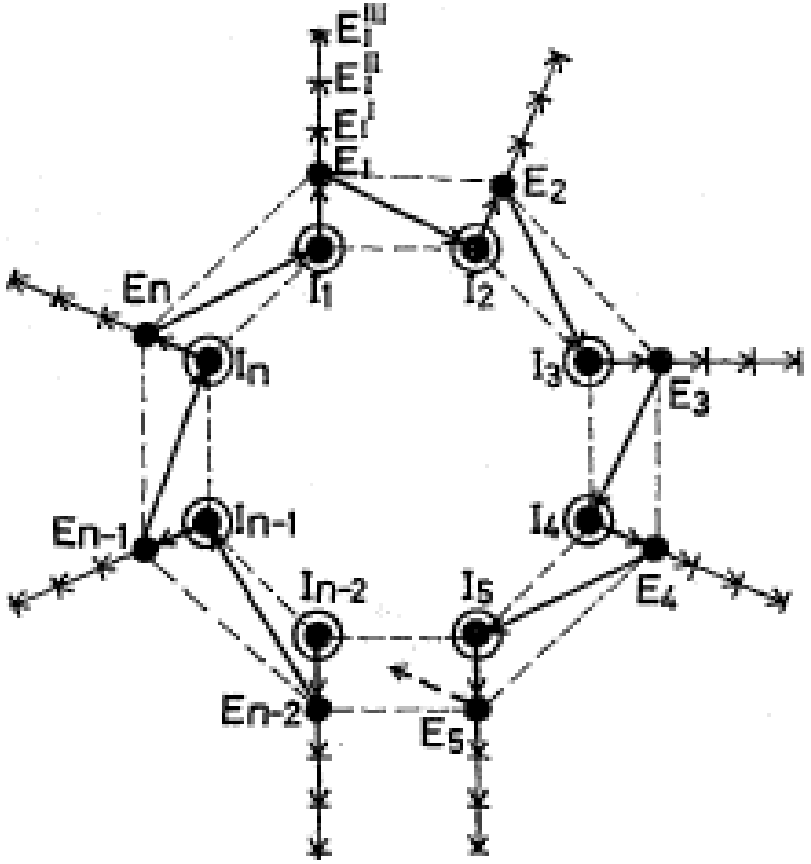


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

5) The cyclic coupling will provide an information capacity which is adapted to the requirements of the system

A Self-reproductive Hyper-Cycle

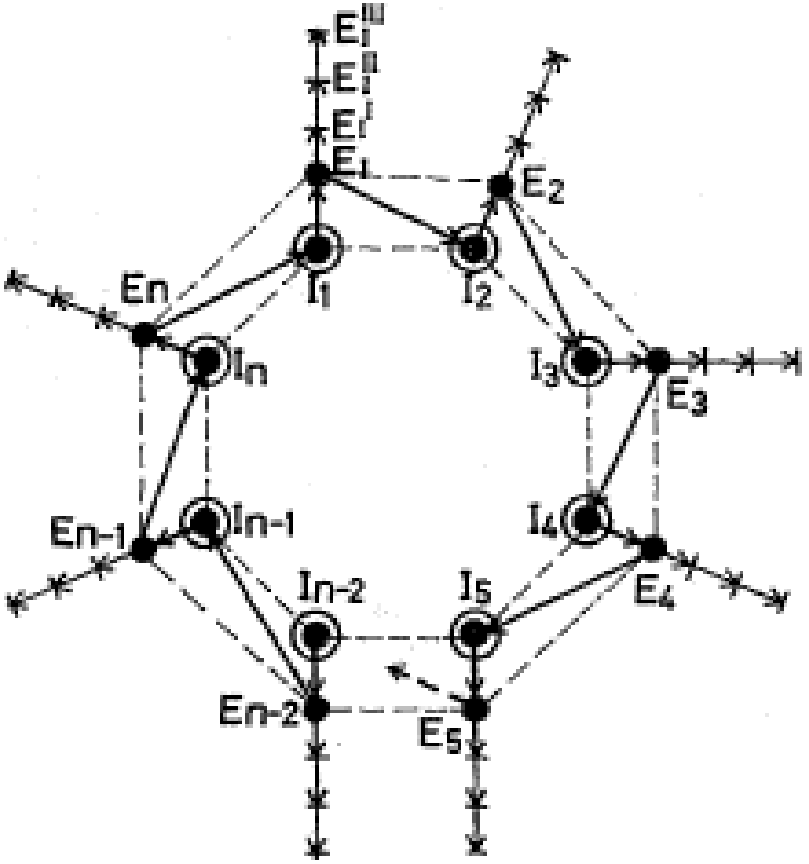


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

6) Genotypic mutations can be used to evolve the system

A Self-reproductive Hyper-Cycle

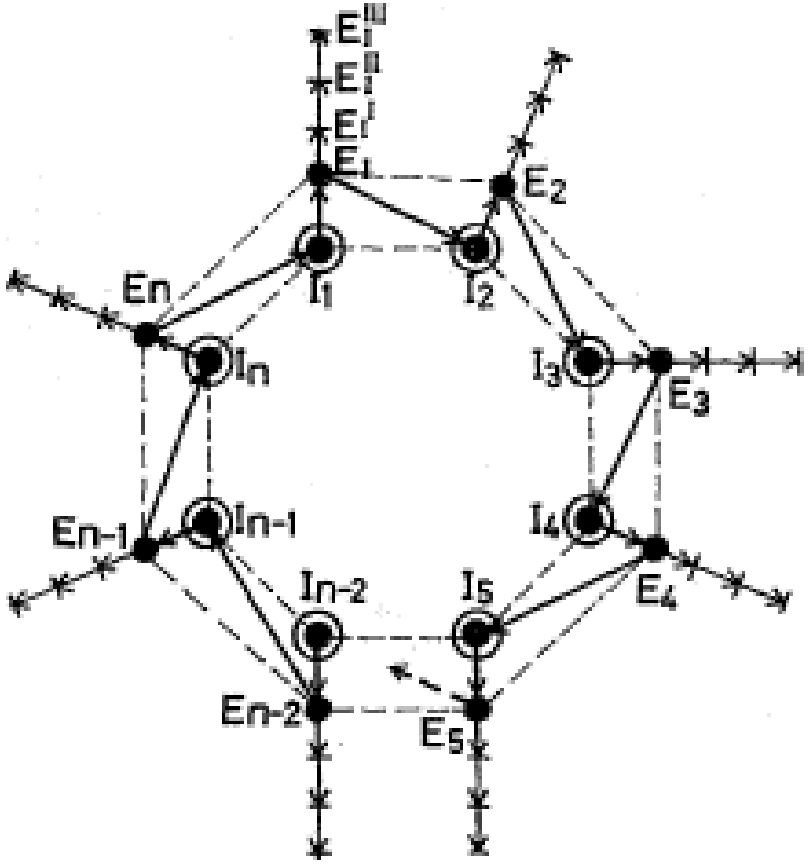


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

7) The system selects against parasitic branches if they have selective values smaller than that of the members of the cycle

A Self-reproductive Hyper-Cycle

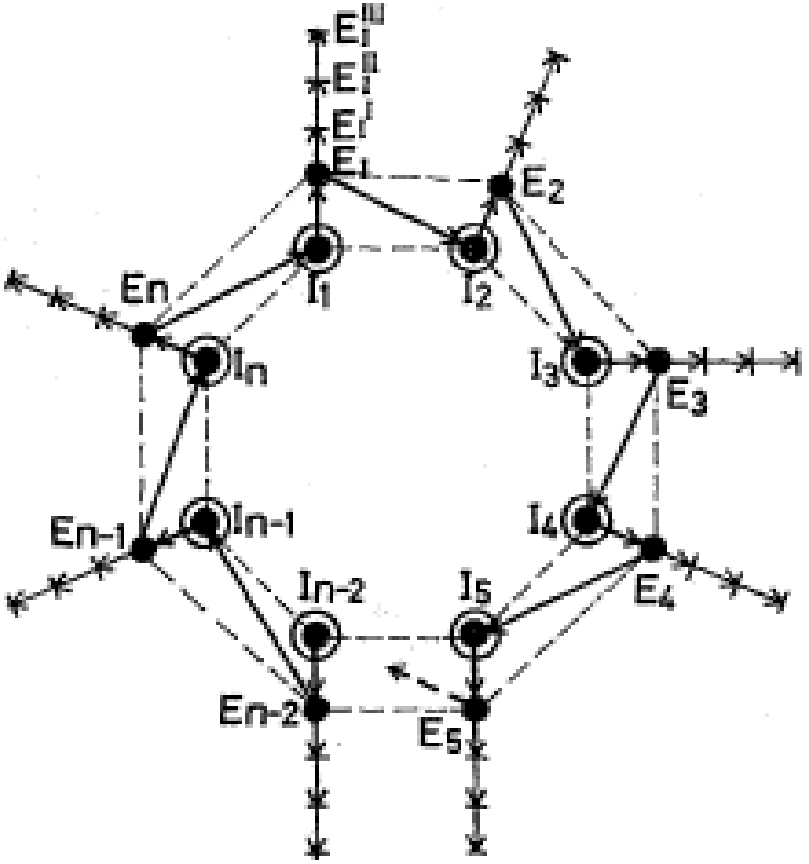


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

8) The only type of branch that can co-exist with the cycle is a branch whose selective value exactly matches that of the cycle

A Self-reproductive Hyper-Cycle

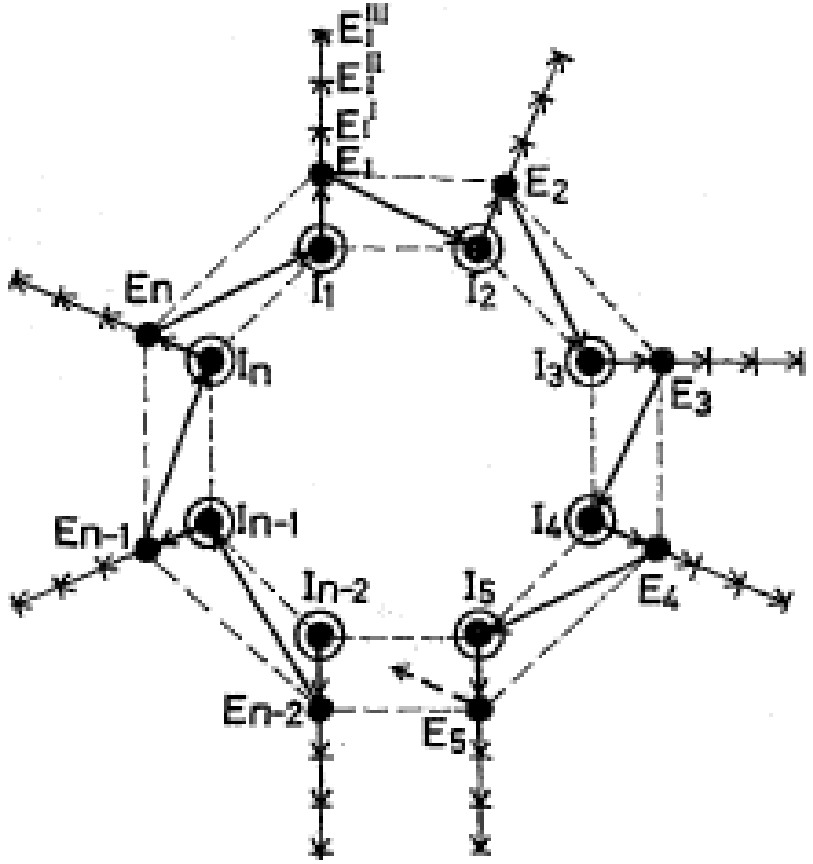


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

9) Only compartmentalized systems can utilize functional branches (brought about by mutations) exclusively to their own advantage (and thus also allow evolution of the branches)

A Self-reproductive Hyper-Cycle

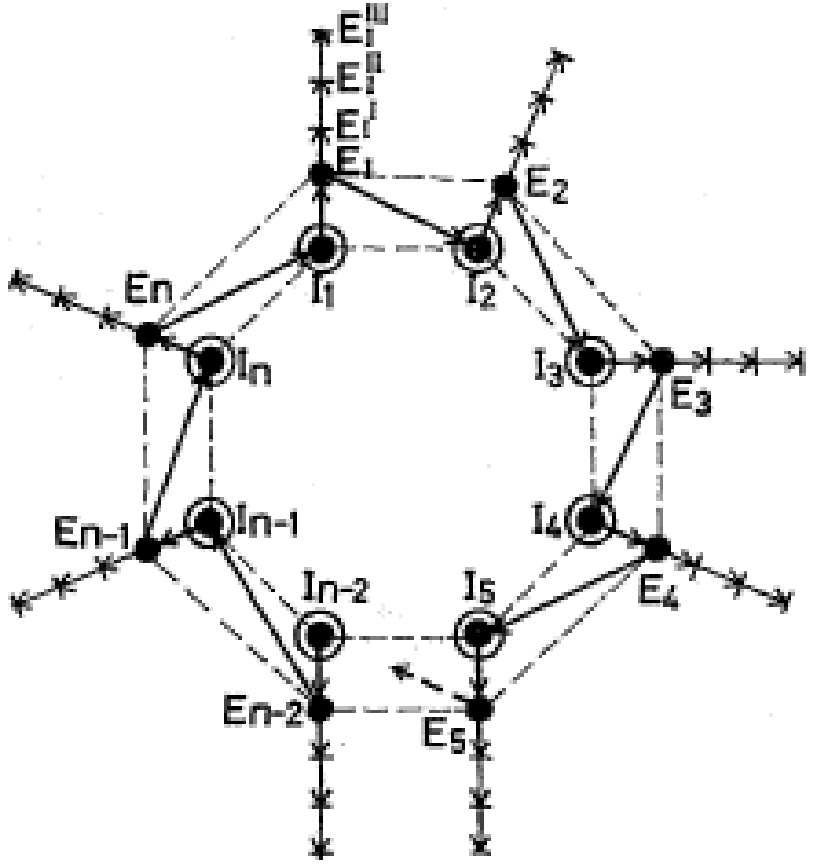


Fig. 15. The selfinstructive catalytic hypercycle

The model consists of:

I_i: nucleotide sequences of limited chain length (complementary single strands of RNA)
E_i: catalytically active polypeptide chains

It is important that this is a closed cycle so that different I_k do not compete and select against each other

10 Properties of the model:

10) A system enclosed in a compartment may individualize by linking its code units into a stable chain

VII: Evolution Experiments

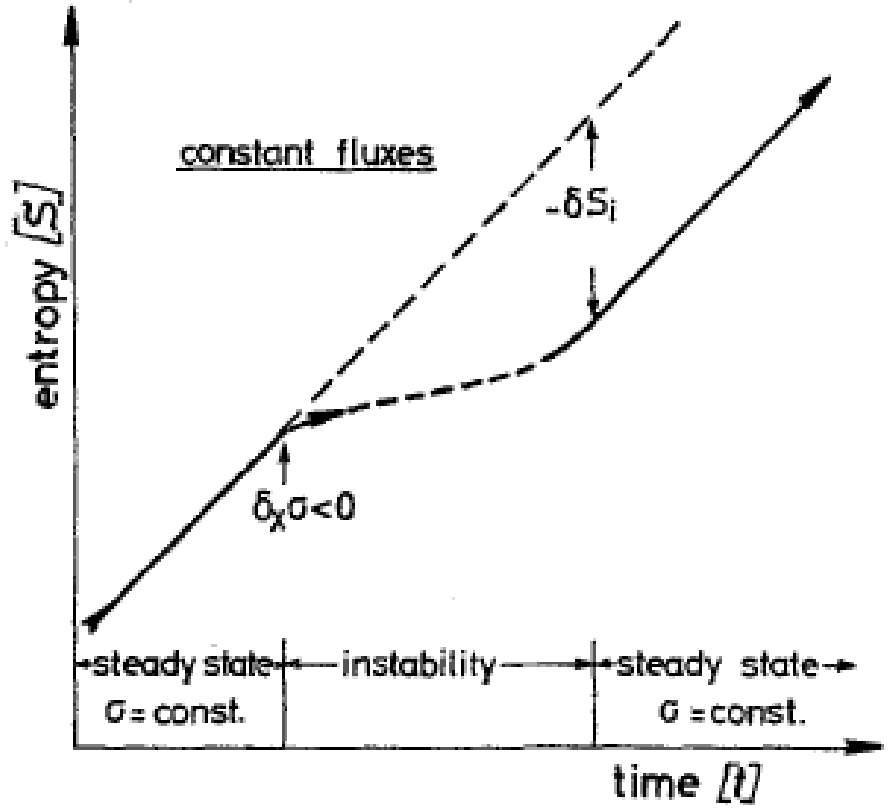
"A theoretical model is only as much as its capacity for experimental testing"

Video

Spiegelman experiments

- QB replicase system: phage QB utilizes a specific replicating enzyme which recognizes exclusively QB RNA
- Infectious viral RNA was reproduced in cell-free media
- Cell-free solution was subject to serial transfer
- The final product contained less than one in 10^{15} of the initial phage template, yet the sample was as infectious as the original one

VIII. Conclusion: Entropy time diagram for a selection process at constant overall flows of digits



- The occurrence of a mutant exhibiting a selective advantage corresponds to a negative fluctuation of entropy production
- This causes an instability/break down of the steady state
- The former master copy dies out and the mutant grows to a dominant level

Conclusion

- The theory explains the general principle of selection and evolution at the molecular level, based on a stability criterion of the thermodynamic theory of steady states
- Evolution appears to be an inevitable event, given the presence of certain matter with specified autocatalytic properties and a finite amount of free energy flow necessary to compensate for the steady state production of entropy
- The theory may explain how to construct simple molecular models representing possible precursors of living cells
- “The fact that ‘selection’ and ‘evolution’ - in a certain analogy to equilibrium in thermodynamics - can be characterized by extremum principles allows a physical foundation and a quantitative formulation of Darwin’s principle”

Moving Forward

"Evolution at the molecular level may be considered a game in which the intelligence of the player is replaced by a selective "instinct" for advantage among randomly occurring events. Therefore game theory, as introduced by John von Neumann, which in recent years has been developed to a high level of sophistication, is the key to any further generalization of evolution theory"

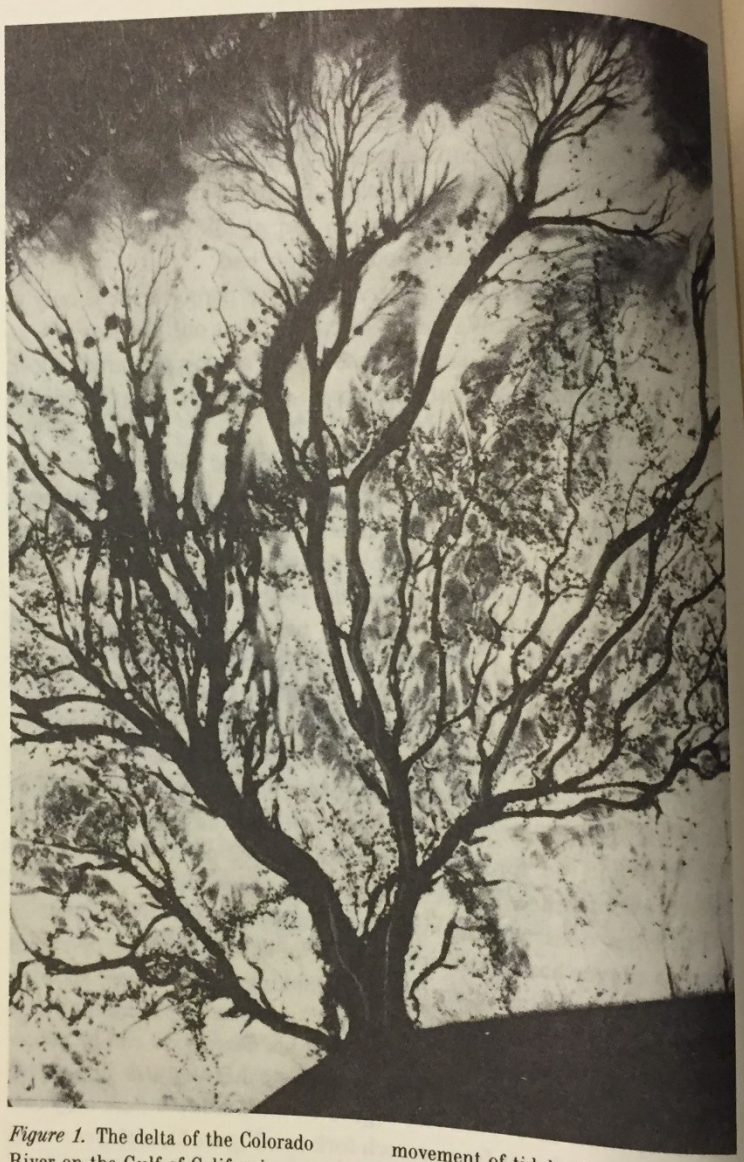


Figure 1. The delta of the Colorado River on the Gulf of California resembles a huge "decision tree." The innumerable branches in the coastal area are caused by the movement of tidal ebb and flow. (Scale 1:20,000. This illustration is used with the permission of the Aero Service Corporation, Philadelphia, Pa.)

Table 1. BEAD GAME "NIM"

A random but not too small number of beads is arbitrarily divided into several groups and placed on a board in separate rows (see Figure 3). The players take turns at removing the beads from the board. At each turn, they may take one or more beads from any one row; or, indeed, they may take a whole row. Whoever removes the last bead is the winner. Each player therefore tries to create a situation that forces his opponent to remove the next-to-last bead or row of beads from the board.

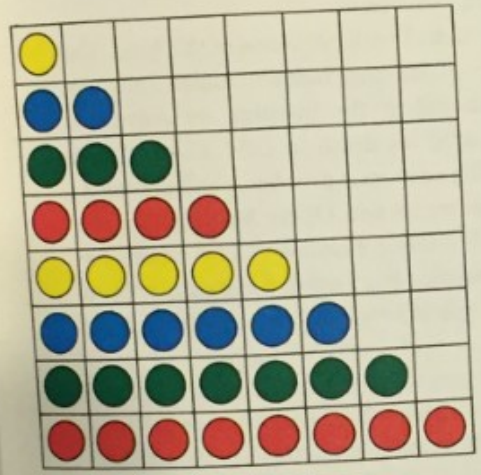


Figure 3. "Nim." One possible layout for a game of "Nim" is illustrated here. Eight groups of colored beads are divided into eight horizontal rows. At each turn, a player may take as many beads as he likes from the board, provided that he takes them from one row only.

two-person, zero-sum game with perfect information and optimal strategy. Translated into everyday language, this forbidding description means simply:

1. that the game is played by two players,
2. that it ends after a finite number of moves have been made,
3. that there are always a winner and a loser, i.e., that the sum of what is won and what is lost always amounts to zero, and
4. that there is one clear strategy that insures victory to the player who makes the first move, regardless of what his opponent may do.

Laws of the Game (Eigen and Winkler)