The emergence of coding specificity at the dawn of life

Peter R Wills

Collaborators: Kay Nieselt, John McCaskill, Peter Stadler, Mohsen Khorsid, Alexei Drummond, Sidney Markowitz, David Bryant, Remco Bouckaert

Overview

1. Biology based on codes.

2. The biological meaning of genetic information: The two classes of aminoacyl-tRNA synthetases are a palimpsest of a primordial binary code that progressively diversified.

3. Systems of coding, reproduction and self-construction.

PART 1

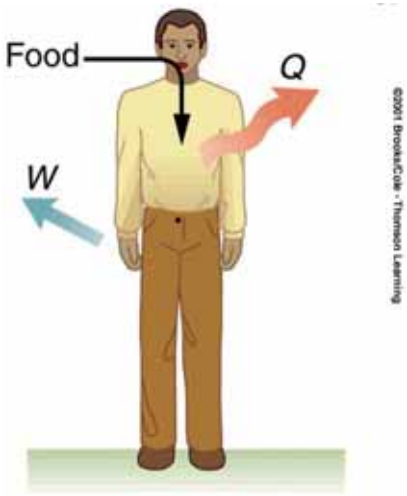
Biology based on codes

- What is life?
- What are codes?

What is life?

Schrödinger (1944): the presence of life in the physical universe can only be explained in terms of:

– dissipation: consumption of "negentropy" to stave of inevitably disordering effects of temperature

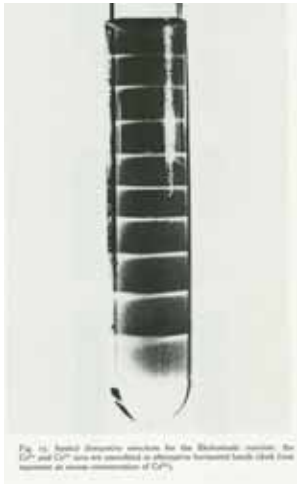


– transmission of heritable information in some physical form, dubbed an "aperiodic crystal"

What is life?

Beyond Schrödinger …

– Prigogine: beyond the threshold of stability, dissipation can *create* thermodynamic order



– Eigen: beyond the error threshold, information can be preserved as a result of Darwinian selection

$$\nu_{\max} \approx \frac{\ln \sigma}{1-\bar{q}}$$

Living systems

- complex molecular components, far-from equilibrium, dynamic

- structurally specific molecular components recurrently synthesized

Living systems

- complex molecular components, far-from equilibrium, dynamic
- structurally specific molecular components recurrently synthesized
- *store of molecular information whose meaning is internally defined*

What is a code?

Generally, abstractly:

(set of)
SIGNS

conventions; rules; mapping

(set of)
MEANINGS

A transformation (of information) between symbols from two alphabets (which may be identical, “copying”).

What is a code?

In biology, in the material world of molecules:

ORGANIC SIGNS → BIOLOGICAL MEANINGS

"WORLD 1" molecular moieties → "WORLD 2" molecular moieties

"general" molecular mechanism

adaptors

Reductionist criticism

Are the biochemical interactions involved in coding qualitatively different from all the other molecular interactions?

Doesn't the "blindness" and clear physicality of the genetic code short-circuit any argument that this or any other code has any kind of "fundamentally different" nature?

If so, it is specious to debate whether the higher-level "codes" are really codes or just enzyme-substrate interactions of a rather general kind.

My response (1)

Information (an ordered set of choices) can become an embodied "thing", able to be the *cause* of events (effects) in a specialized physical context, when there is a stable general mechanism of some kind whose repeated operation produces different results (outcomes) for a range of initial conditions (inputs), i.e., when there is a code.

The ordering of the "choices" can be spatial (bases along a DNA heteropolymer) or temporal (changes in messenger concentrations).

My response (2)

A stably operating general mechanism can serve as a platform for an organic code when the piecewise mapping from inputs to outputs is of significance in relation to the state and orderly dynamics (the *existence* even) of a larger system that supports it.

In these terms, an organic code can only exist in a system that is, broadly speaking, *autocatalytic*, because the system must be able to maintain synthesis of the molecular components of the coding machinery.

Example: the genetic code

The stably operating general mechanism is protein synthesis, collinear with mRNA.

This is of significance in relation to the state and orderly dynamics (the *existence* even) of a cell because it promotes catalysis and its control.

The mapping from mRNA sequence inputs to protein sequence outputs is piecewise: codons to amino acids.

Example: the genetic code

Protein synthesis is, broadly speaking, *autocatalytic*, because many of the molecular components of the protein synthetic machinery are themselves proteins.

Coding is autocatalytic, because the individual, independent coding tools, the amino acyl-tRNA synthetases, are proteins.

Note: these tools catalyse formation of the passive "adaptors" (charged tRNAs) through which coding is maintained during peptidyl transfer.

THE question concerning biology

How can information and meaning, the egg and the chicken, come into mutual existence? [A DNA heteropolymer can only be said to contain sequence information in relation to a sequence-sensitive process.]

How can more and more complex versions of information and meaning continue to emerge, when each new more complex version needs all of its more complex parts to operate and sustain itself?

Answering THE question

How can information and meaning, the egg and the chicken, come into mutual existence?

How can more and more complex versions continue to emerge, when each new more complex version needs all of its more complex parts, more precisely specified components, to operate and sustain itself?

ANSWER: *through progressive dynamical bifurcations in which the dissipative flow of free energy is sequestered into more narrowly defined, precisely specified biochemical channels.*

EXAMPLE: *evolution of the amino acyl-tRNA synthetases, i.e., the refinement of the genetic code*

Answering THE question

ANSWER: *through progressive dynamical bifurcations in which the dissipative flow of free energy is sequestered into more narrowly defined, precisely specified biochemical channels.*

The terms "*more narrowly defined*" and "*more precisely specified*" imply an *increase in the quantity of information* involved – for example,

from choosing an amino acid according to a classification into $N = 2$ simple classes

to choosing an amino acid according to the canonical classification into $N = 20$ distinguishable types

Measuring information

Shannon's formula $H = -\sum_{i=1}^{N} p_i \log_2 p_i$

provides a convenient measure of the amount of information. But its use requires the a priori specification of the class of *N* possibilities (outcomes) over which the probabilities p_i are summed.

An organic code requires the operation of a general molecular mechanism, that produces (predominantly) one of *N* different outcomes given different inputs belonging to a general class of entities/conditions *that are in some way otherwise equivalent.*

Specifying information

A body of information, like the sequence of the human genome, can only be described by reciting it (perhaps using an algorithm to do so, if it is appropriately compressed). It is a *nominable* entity (Barbieri, 2013).

The effect of information in biology must be understood in terms of the capability of a nominal entity to cause a very specific outcome (potentially one of v^N possibilities for a sequence of length v chosen from an alphabet of size N).

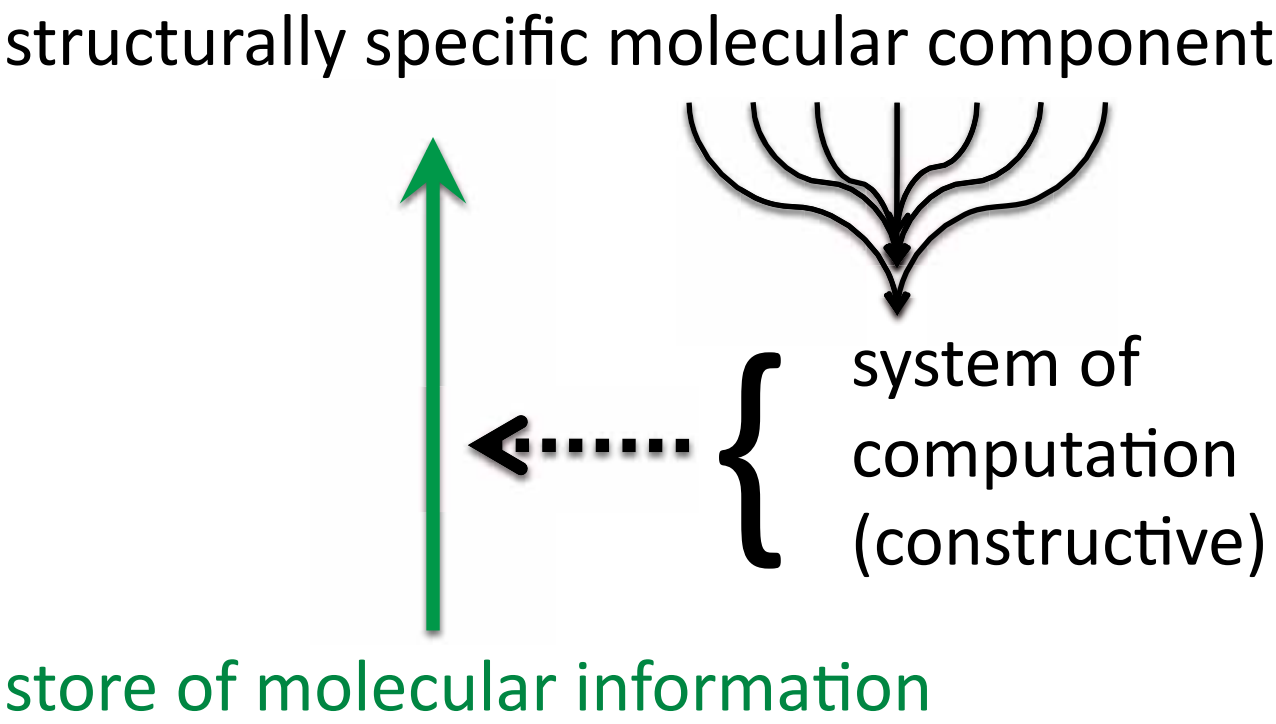
The last nagging question

Why and how do well-defined chemical alphabets, which require their prior specification for them to be of any effect, emerge as classes of (otherwise) equivalent entities, in physical systems? When does natural selection favour *general* solutions over specific, efficacious solutions to problems?

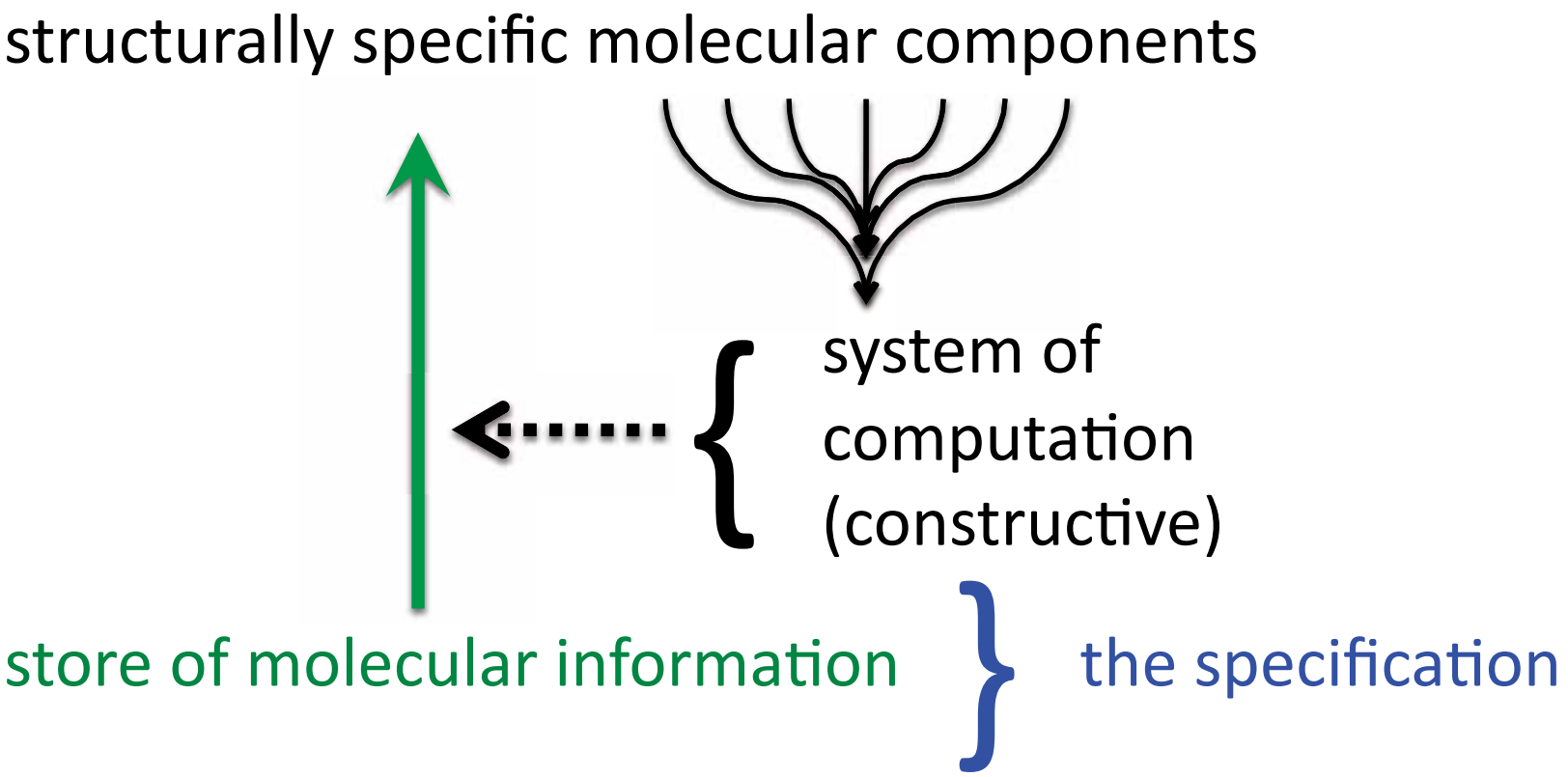
Enough of this abstract theory

LET'S GET REAL

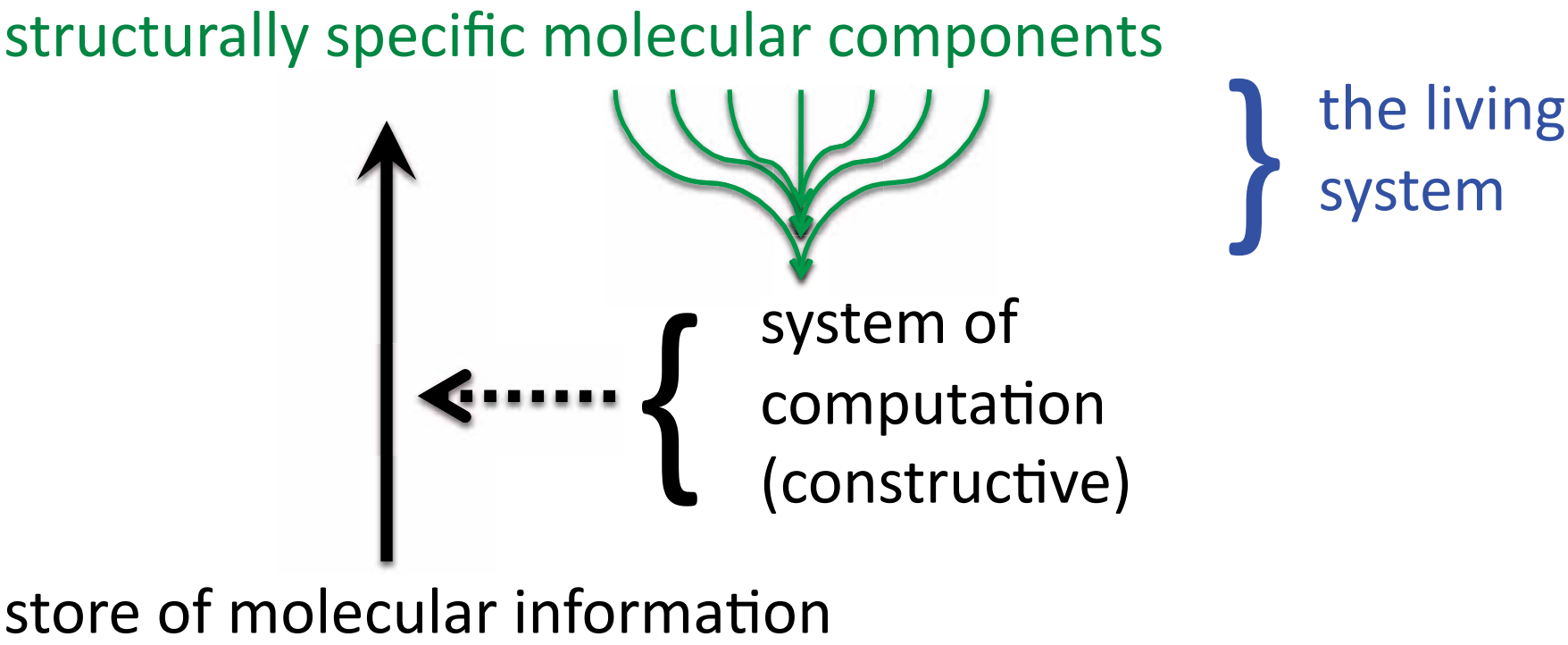
Life requires genetic representation



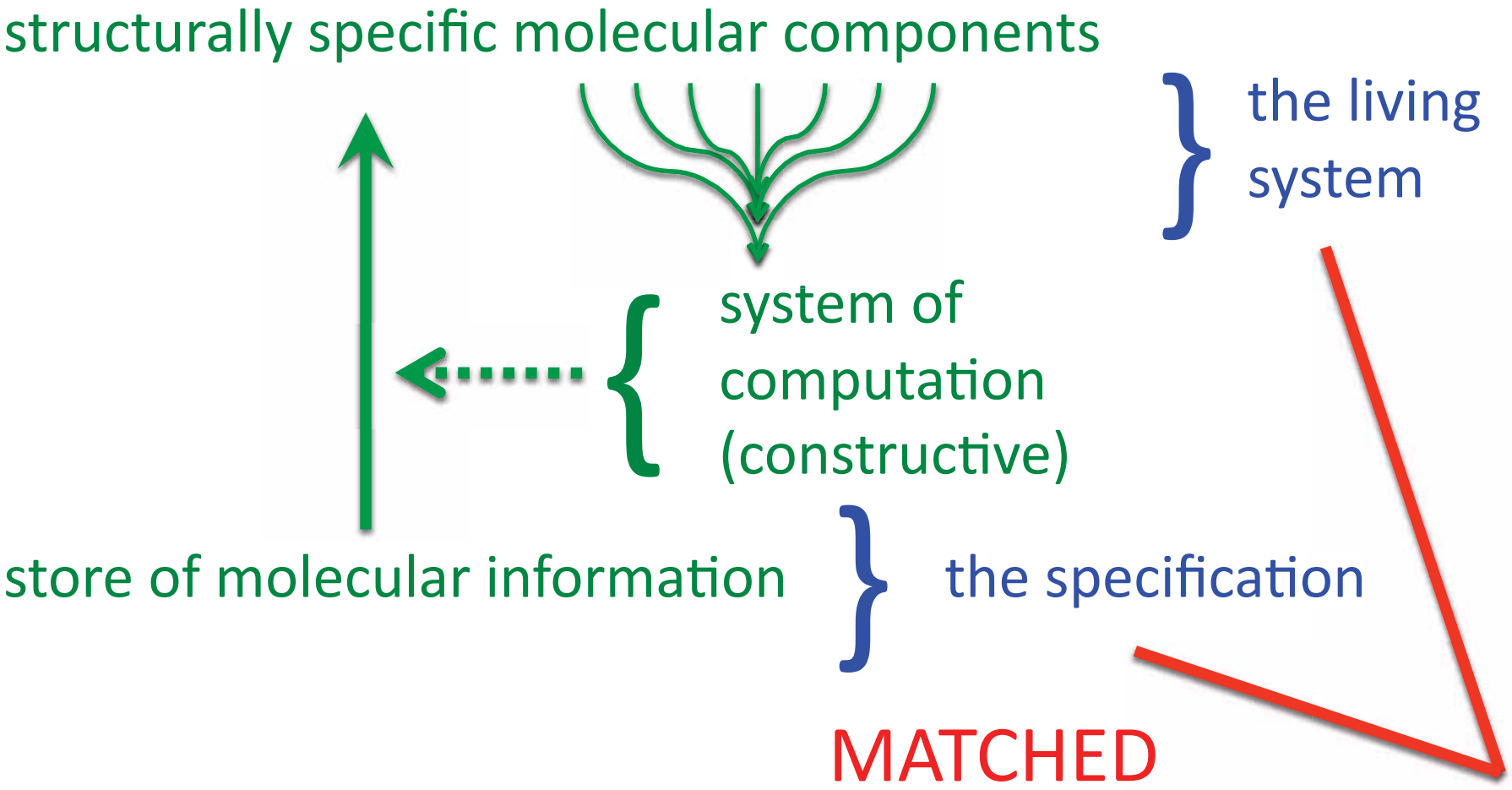
Life requires genetic representation



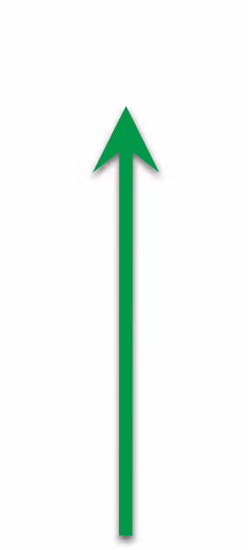
Life requires genetic representation



Life requires genetic representation



Continuity of the genotype

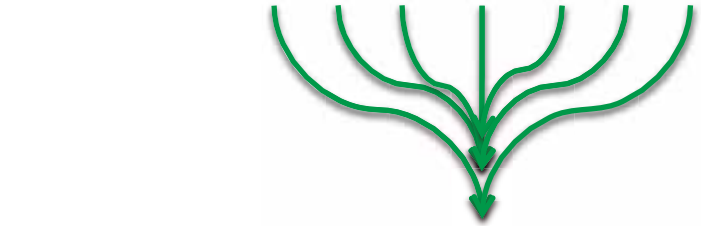


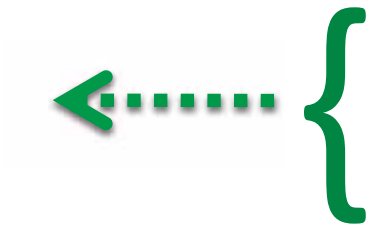
store of molecular information

3.5×10^9 years of replication mutation, and selection

Continuity of the phenotype

structurally specific molecular components





system of computation (constructive)

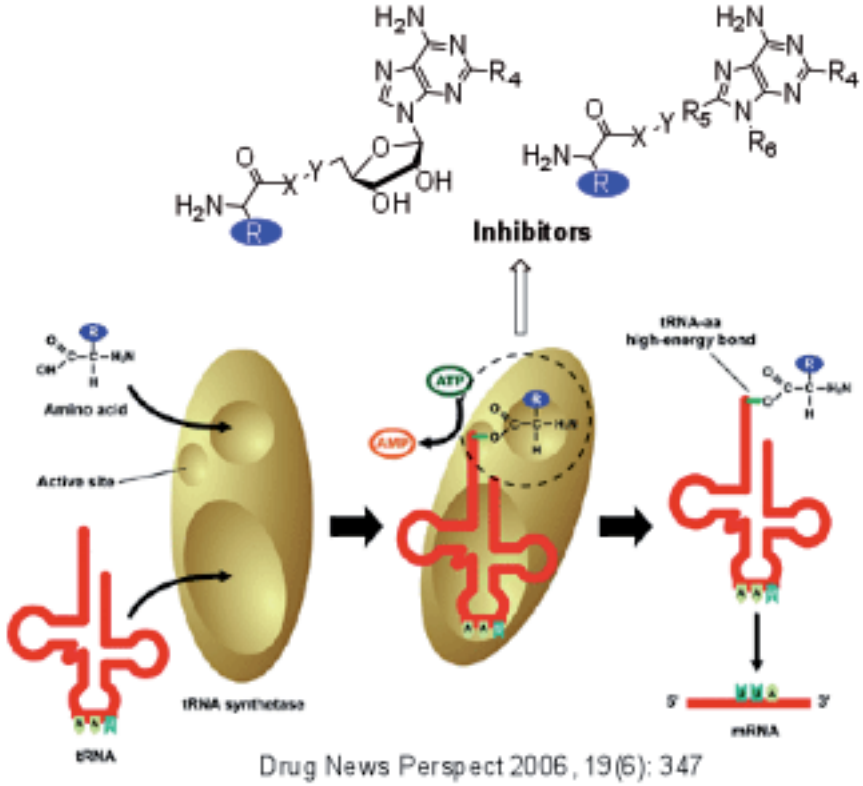
3.5×10^9 years of thermal dissipation, successive symmetry-breaking transitions

PART 2

The AARSs – aminoacyl-tRNA synthetases:

– their central importance for living systems

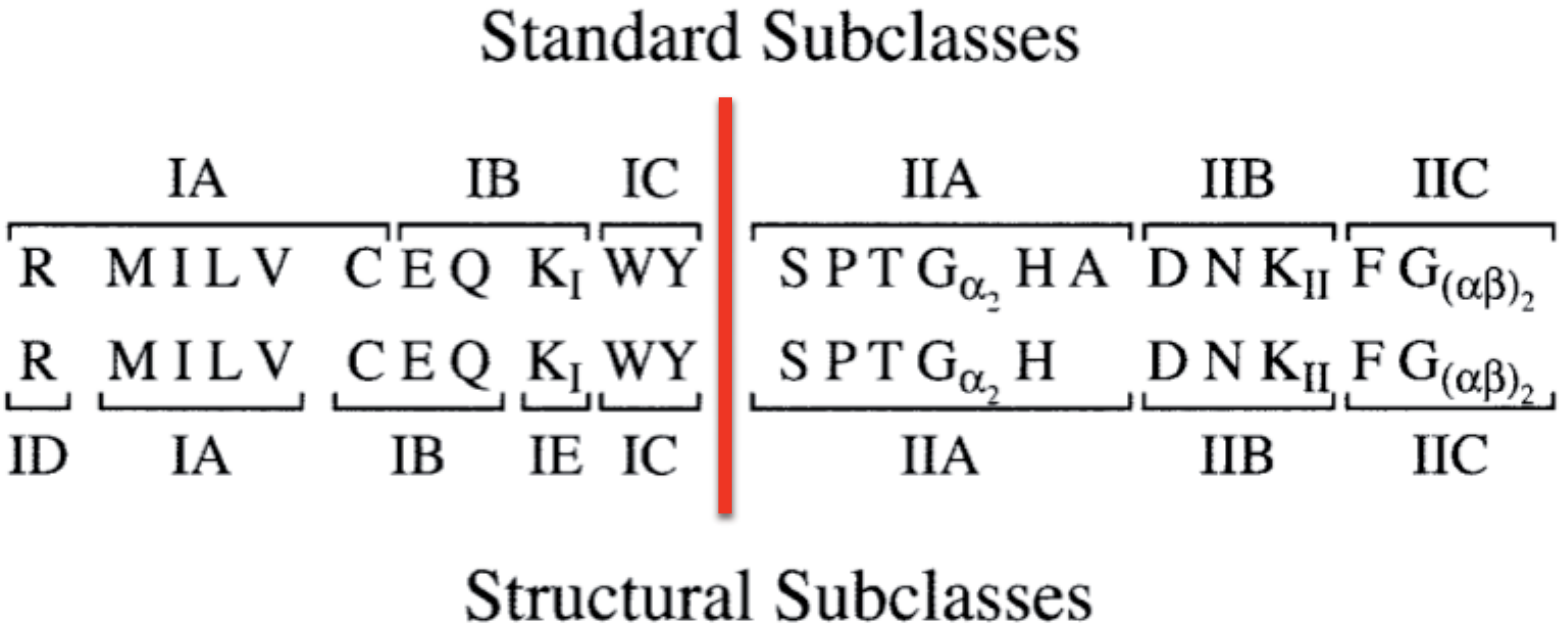
How the genetic code works



Aminoacylation of tRNA

Matching an amino acid with its (anti)codon

Classification of AARSs
The two separate classes, I and II

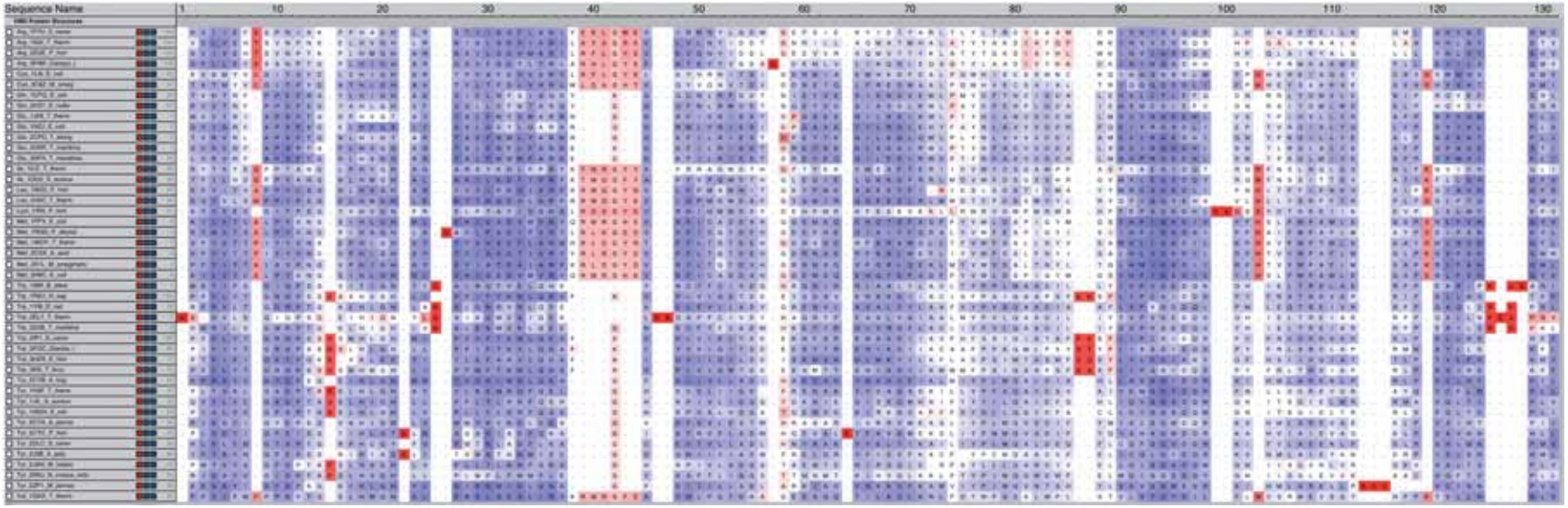


Core Structures of the AARS enzymes

Class I

Class II

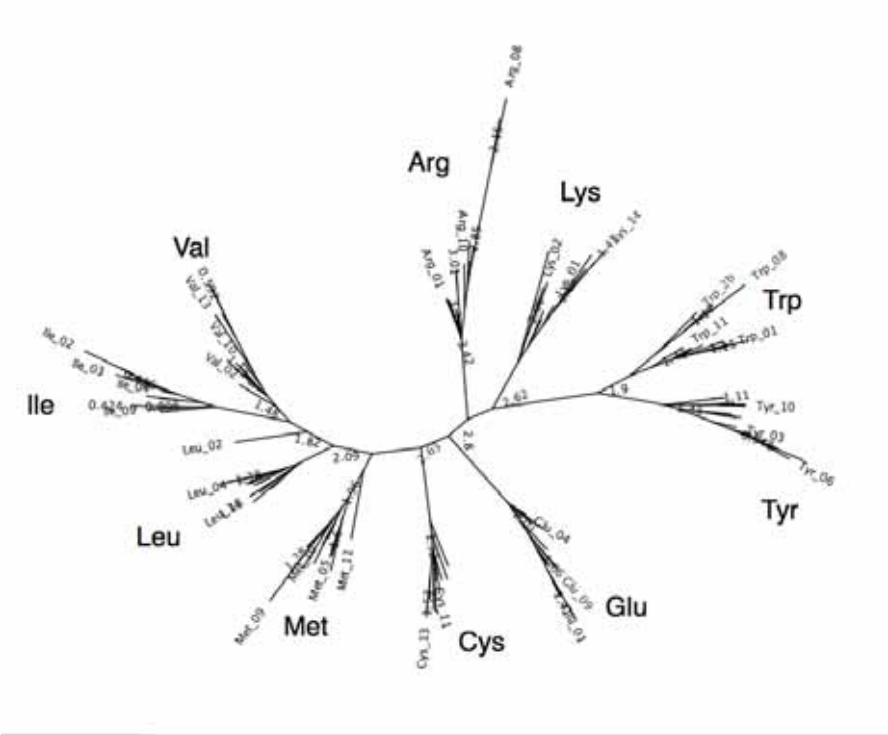
Structural alignment of Class I AARS sequences of all amino acid specificities from species from all kingdoms

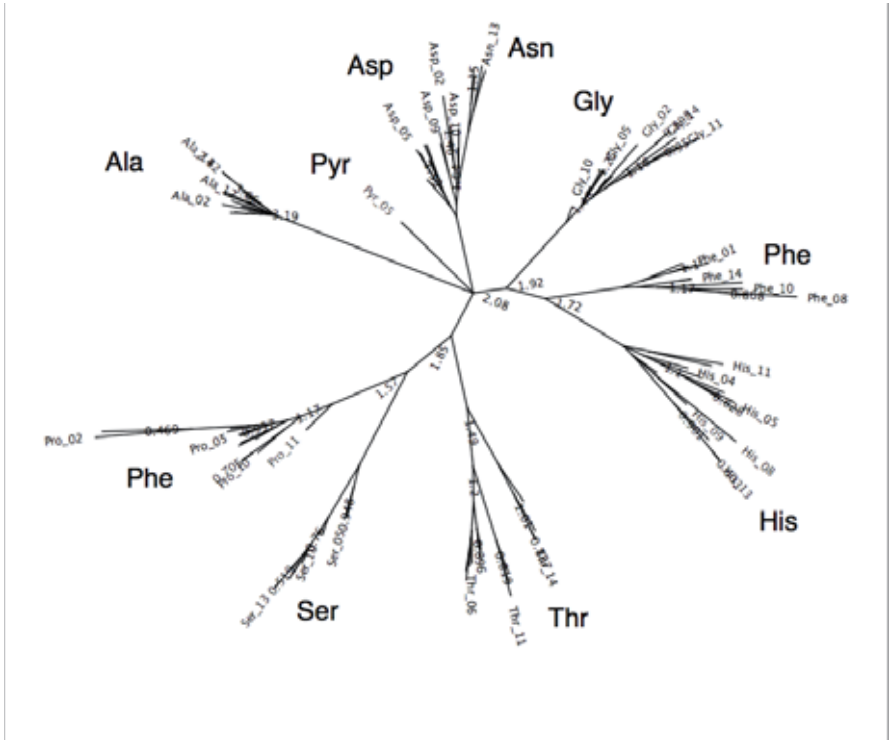


And similarly the Class II structures …

Conventional Phylogeny of the AARSs

for the archaea, drawn as unrooted trees

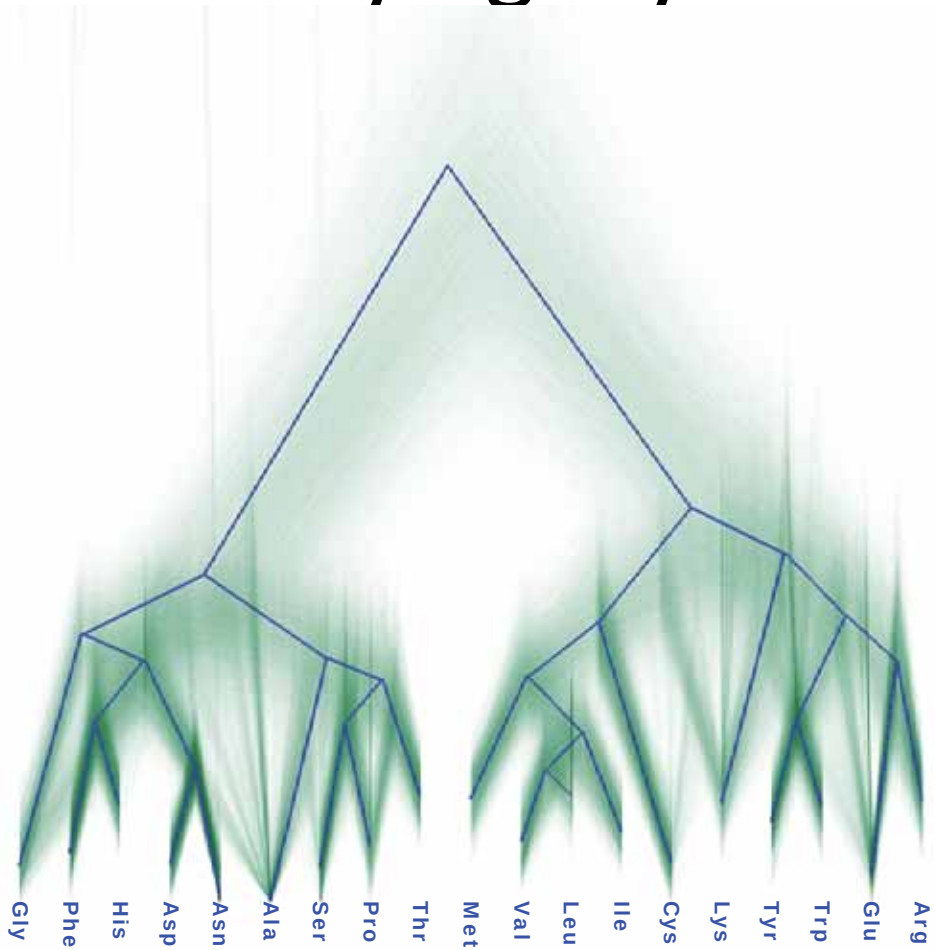




Class I

Class II

Conventional Phylogeny of the AARSs



Class II

Class I

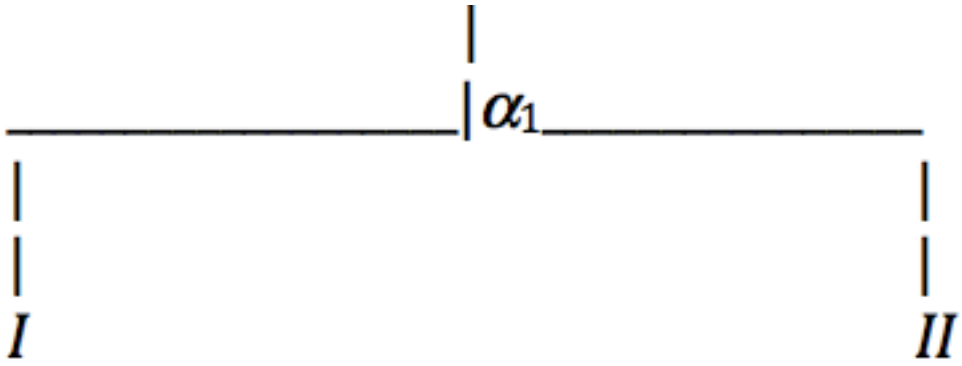
Strange problem

Q Why are there still two solutions – exactly the same two solutions – to one evolutionary problem found in every modern living cell?

Strange problem

Q Why are there still two solutions – exactly the same two solutions – to one evolutionary problem found in every modern living cell?

A? Because the two forms are a palimpsest of a primordial binary code that required a strong separation of forms to maintain the separation of assignment functions $I \rightarrow i$ and $II \rightarrow ii$ without risking either $I \rightarrow ii$ or $II \rightarrow i$

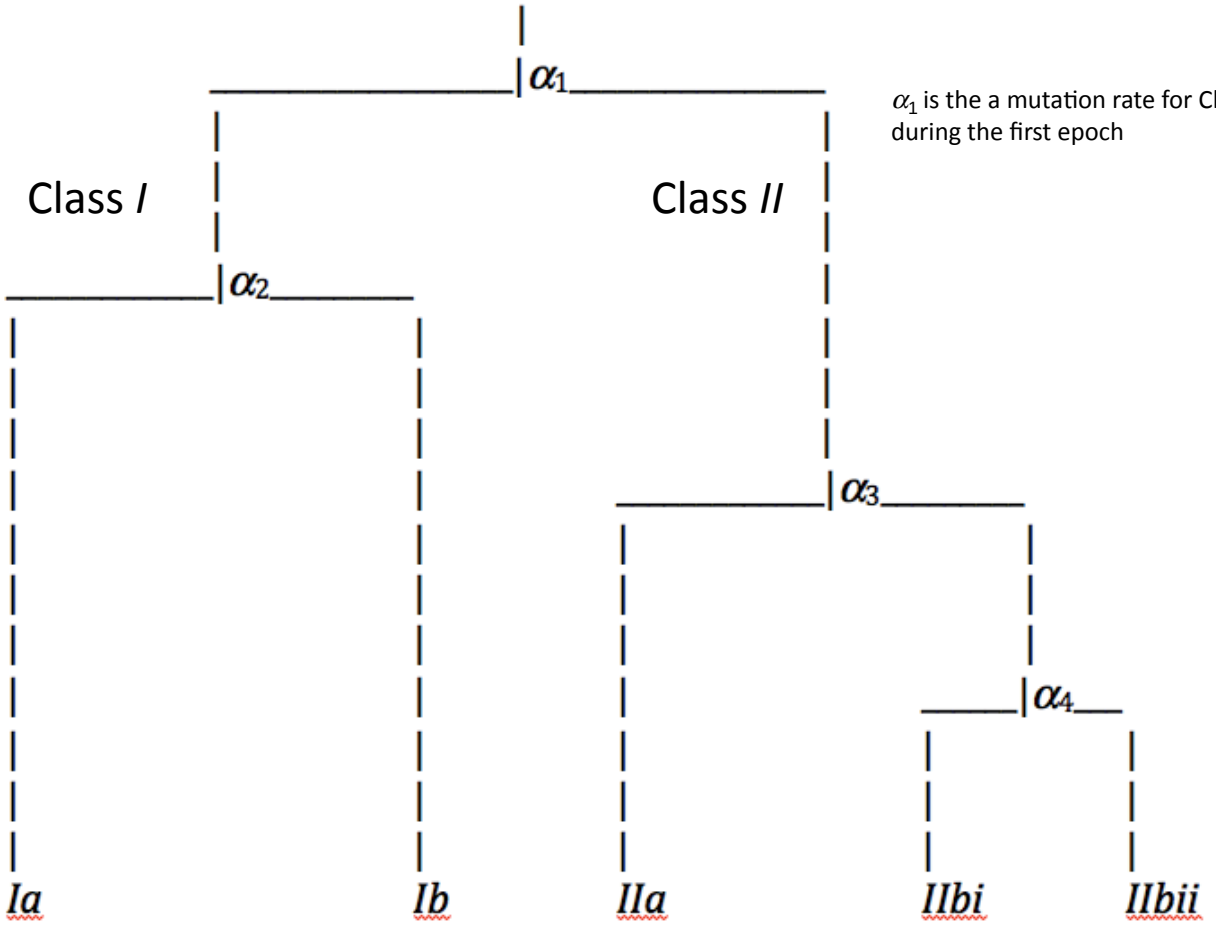


$$R_2 = \begin{pmatrix} I & II \\ - & \alpha_1 \\ \alpha_1 & - \end{pmatrix}$$

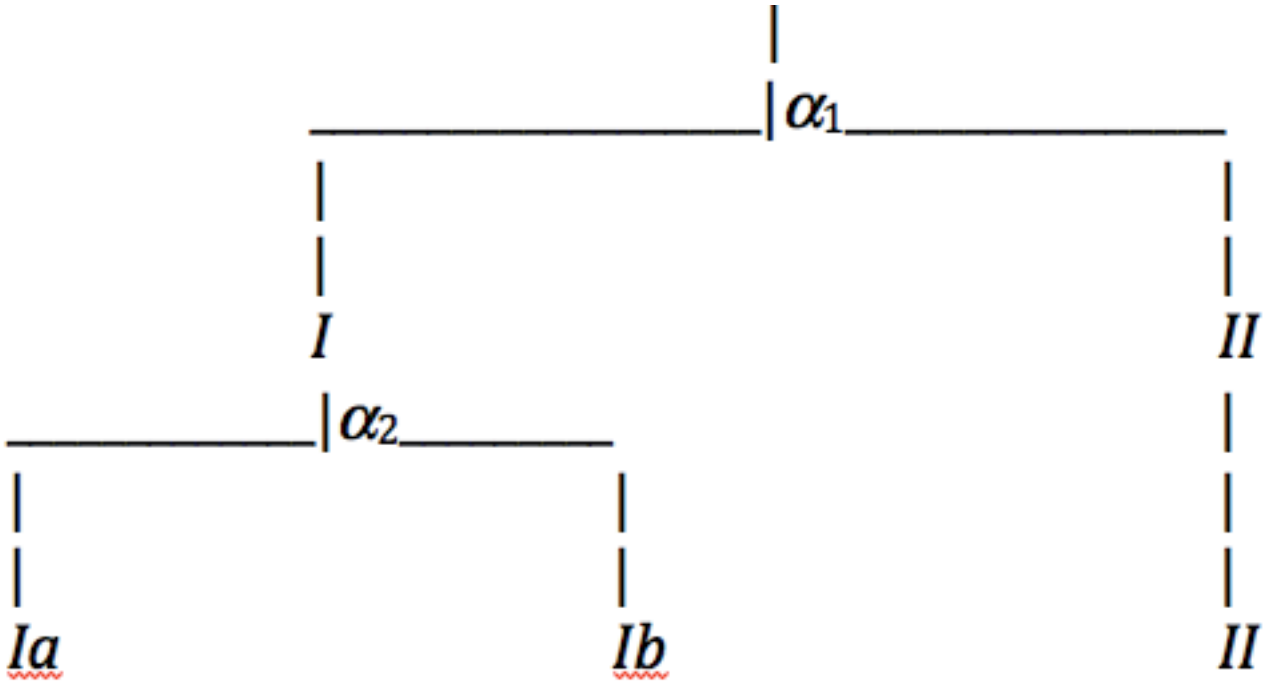
α_1 is the a mutation rate for Class *I* to *II* substitutions during the first epoch

– the era of the primordial binary code

Analysing AARS Phologeny

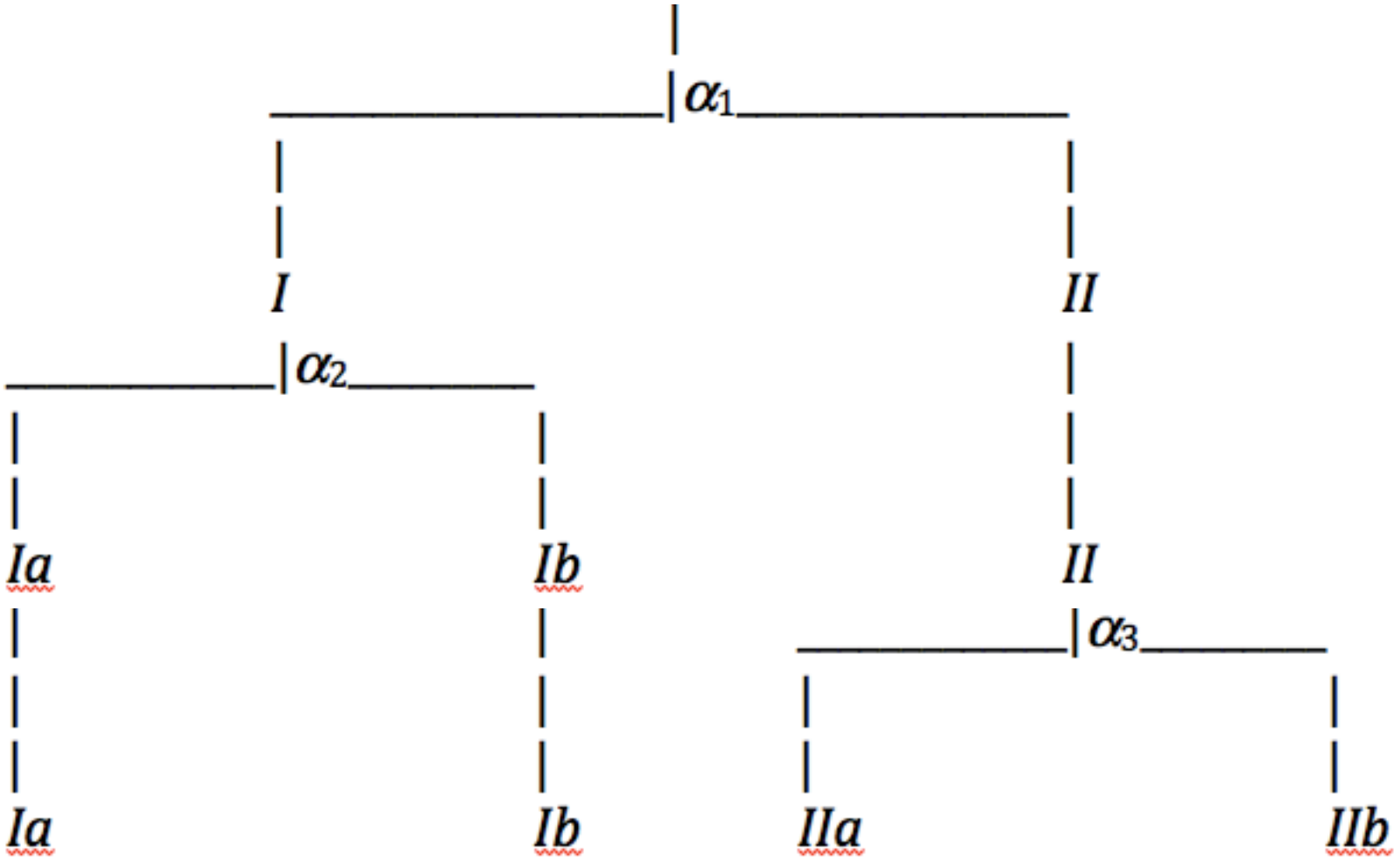


α_1 is the a mutation rate for Class *I* to *II* substitutions during the first epoch



$$R_3 = \begin{pmatrix} Ia & Ib & II \\ - & \alpha_2 & \alpha_1 \\ \alpha_2 & - & \alpha_1 \\ \alpha_1 & \alpha_1 & - \end{pmatrix}$$

α_2 is the a mutation rate for Class *Ia* to *Ib* substitutions during the first epoch



$$R_4 = \begin{pmatrix} Ia & Ib & IIa & IIb \\ - & \alpha_2 & \alpha_1 & \alpha_1 \\ \alpha_2 & - & \alpha_1 & \alpha_1 \\ \alpha_1 & \alpha_1 & - & \alpha_3 \\ \alpha_1 & \alpha_1 & \alpha_3 & - \end{pmatrix}$$

Build up substitution matrix through different aeons

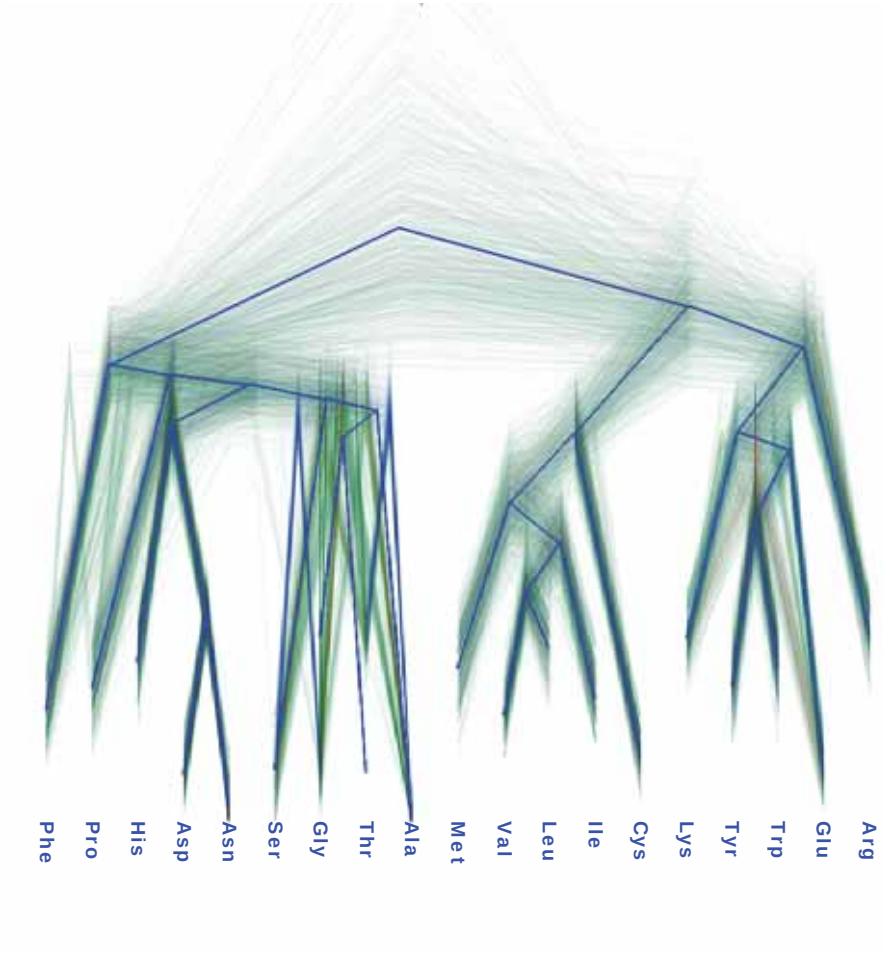
$$R_2 = \begin{pmatrix} I & II \\ - & \alpha_1 \\ \alpha_1 & - \end{pmatrix}$$

$$R_3 = \begin{pmatrix} Ia & Ib & II \\ - & \alpha_2 & \alpha_1 \\ \alpha_2 & - & \alpha_1 \\ \alpha_1 & \alpha_1 & - \end{pmatrix}$$

$$R_4 = \begin{pmatrix} Ia & Ib & IIa & IIb \\ - & \alpha_2 & \alpha_1 & \alpha_1 \\ \alpha_2 & - & \alpha_1 & \alpha_1 \\ \alpha_1 & \alpha_1 & - & \alpha_3 \\ \alpha_1 & \alpha_1 & \alpha_3 & - \end{pmatrix}$$

$$R_5 = \begin{pmatrix} Ia & Ib & IIa & IIbi & IIbii \\ - & \alpha_2 & \alpha_1 & \alpha_1 & \alpha_1 \\ \alpha_2 & - & \alpha_1 & \alpha_1 & \alpha_1 \\ \alpha_1 & \alpha_1 & - & \alpha_3 & \alpha_3 \\ \alpha_1 & \alpha_1 & \alpha_3 & - & \alpha_4 \\ \alpha_1 & \alpha_1 & \alpha_3 & \alpha_4 & - \end{pmatrix}$$

Unconventional Phylogeny of the AARSs



Class II

Class I

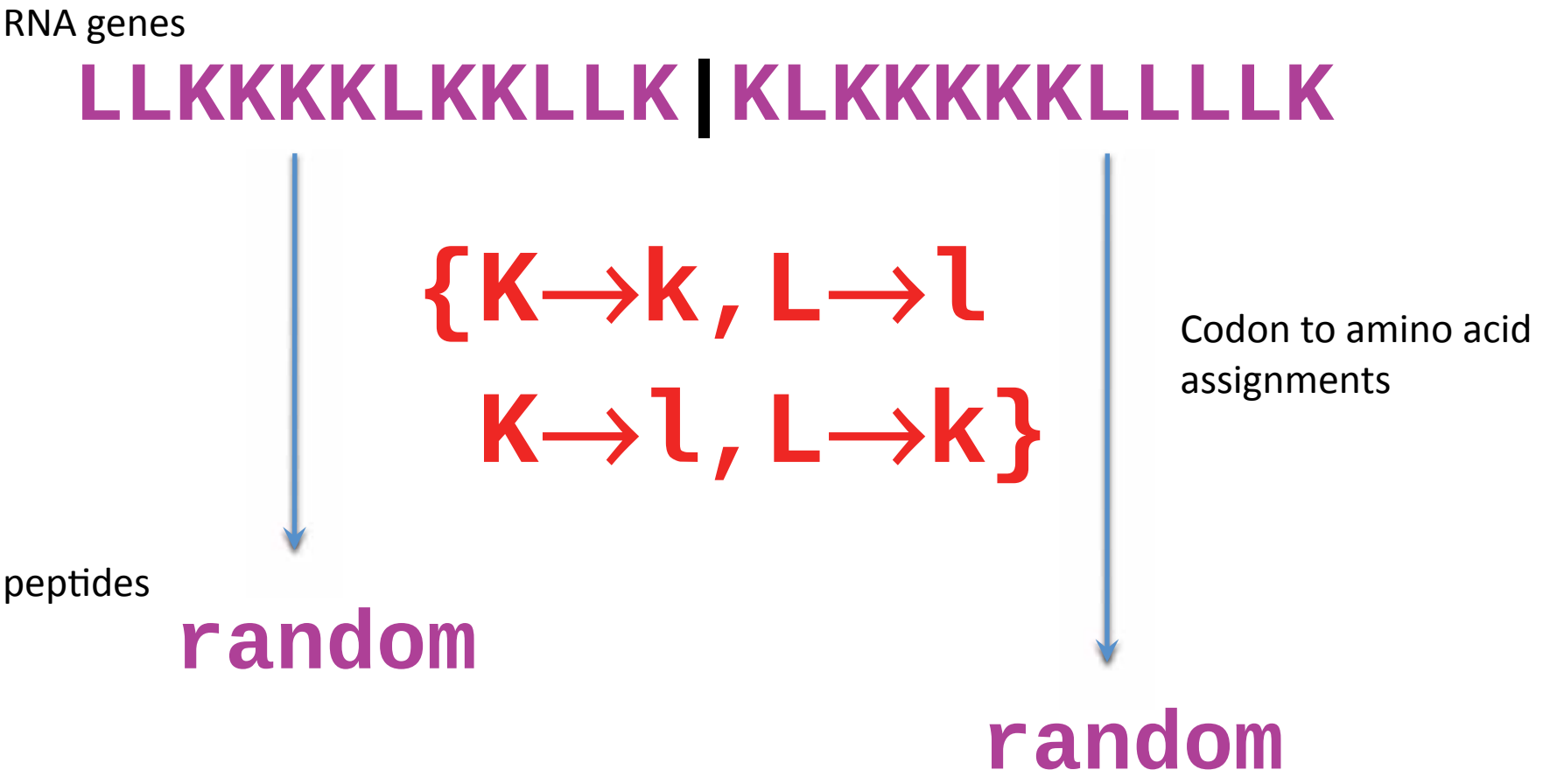
How did this happen?

Not by adaptive mutation!

PART 3

Systems of coding, reproduction and self-construction

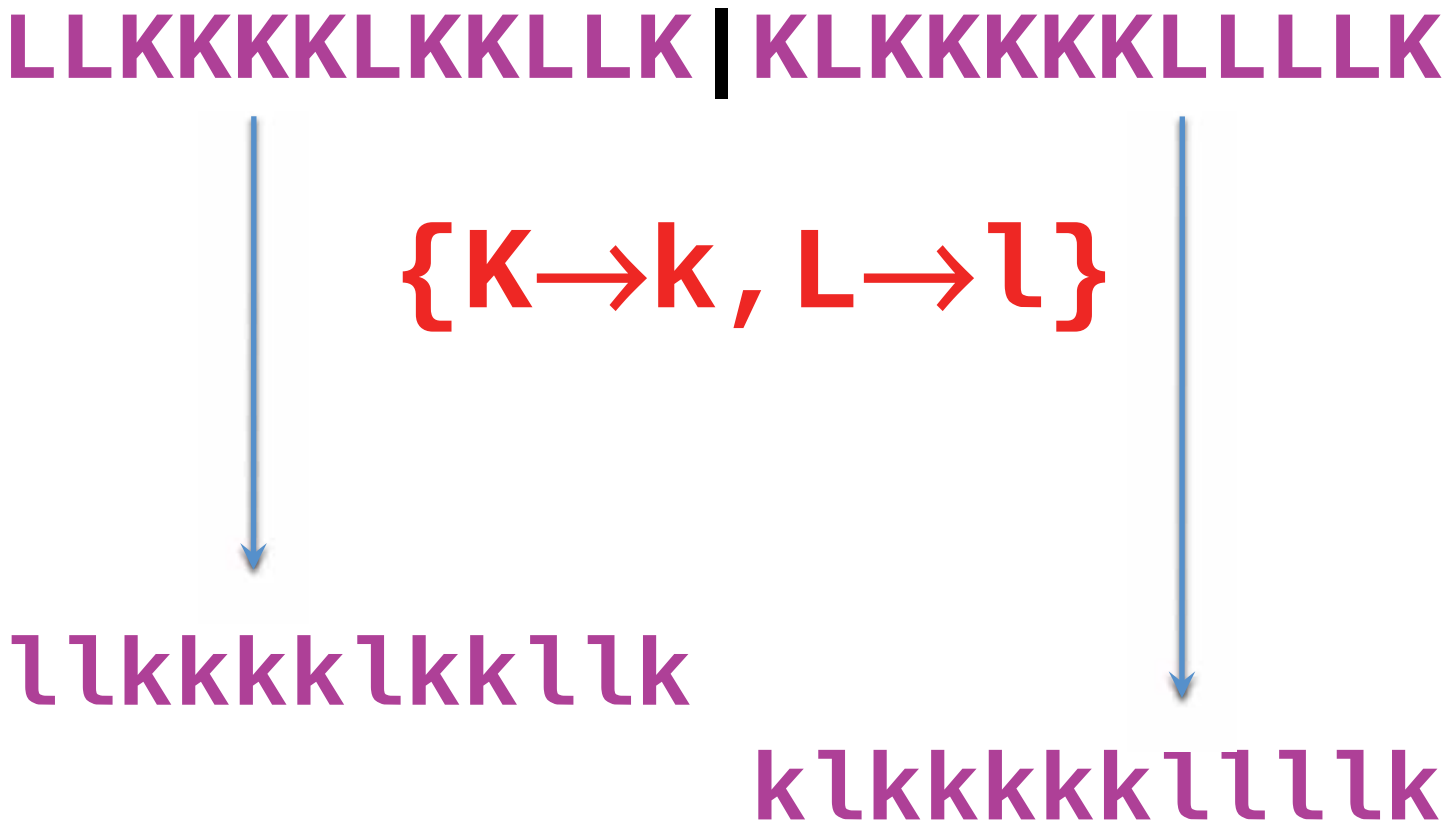
Coding self-organization



Peter Dittrich, slide 16

Equi-probable **X→y** assignments

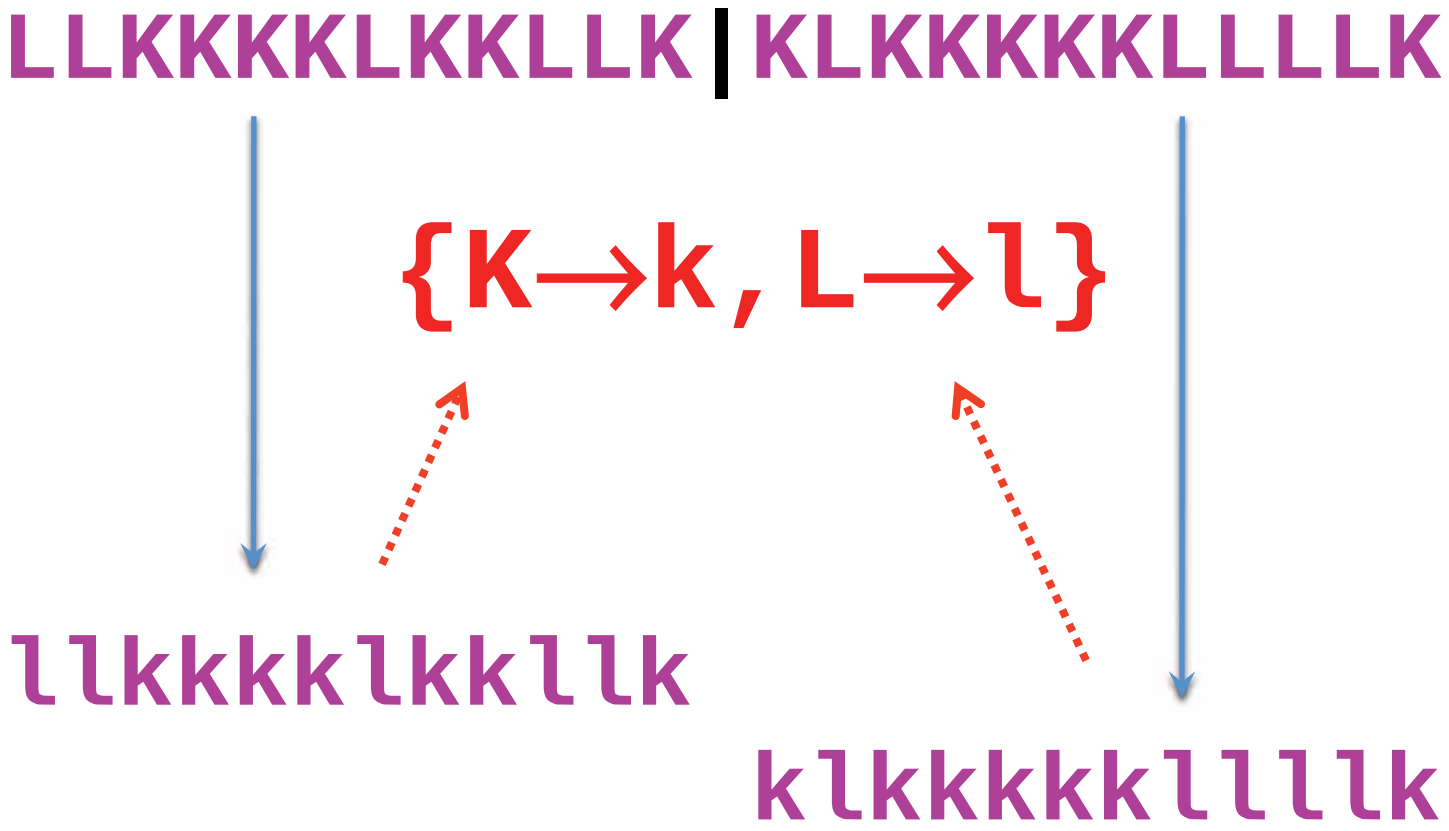
Coding self-organization



Peter Dittrich, slide 16, choice of a “chemistry”

Coded **K→k, L→l** assignments

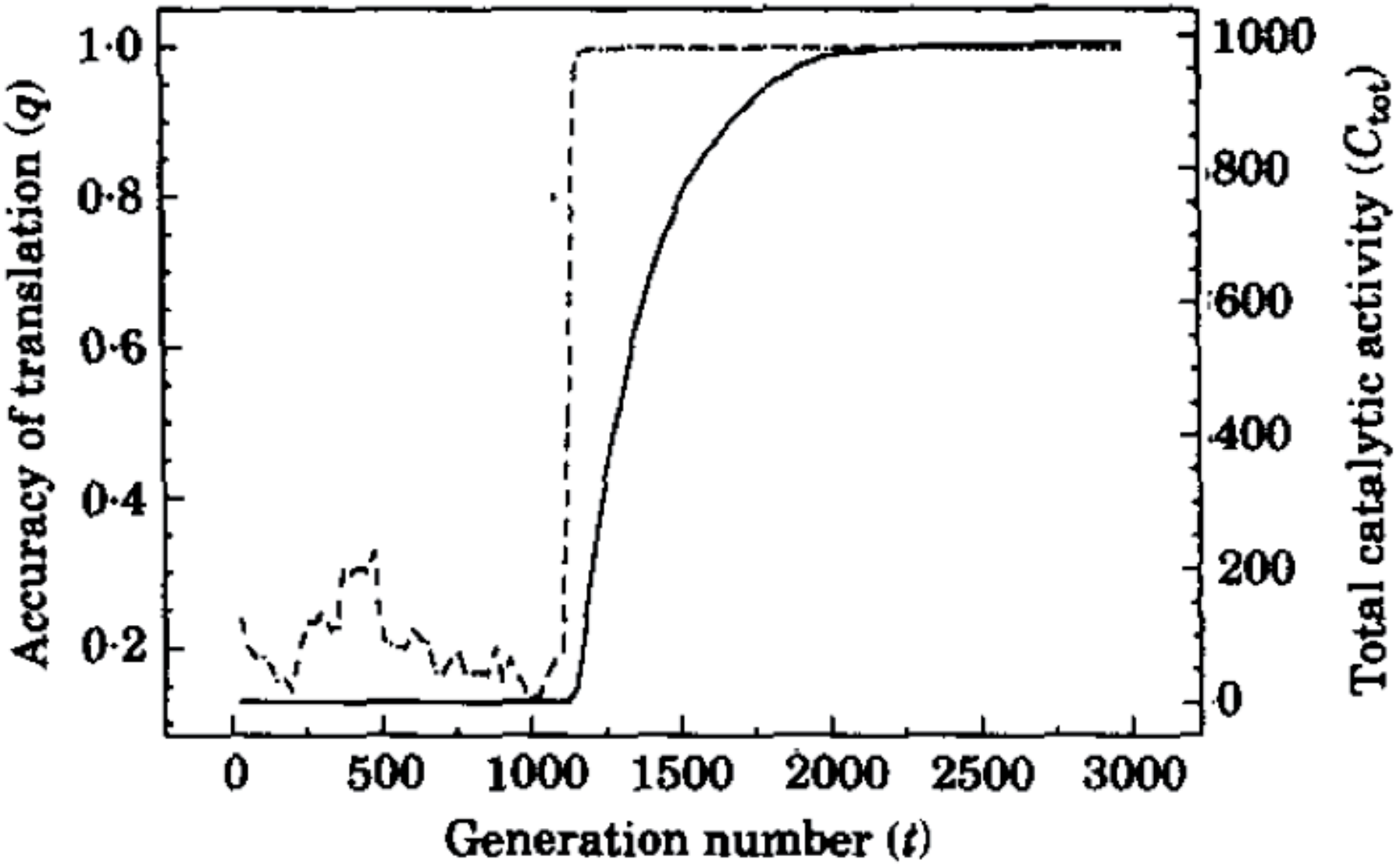
Coding self-organization



Peter Dittrich, slide 16, choice of a "chemistry", but here, *by the system*

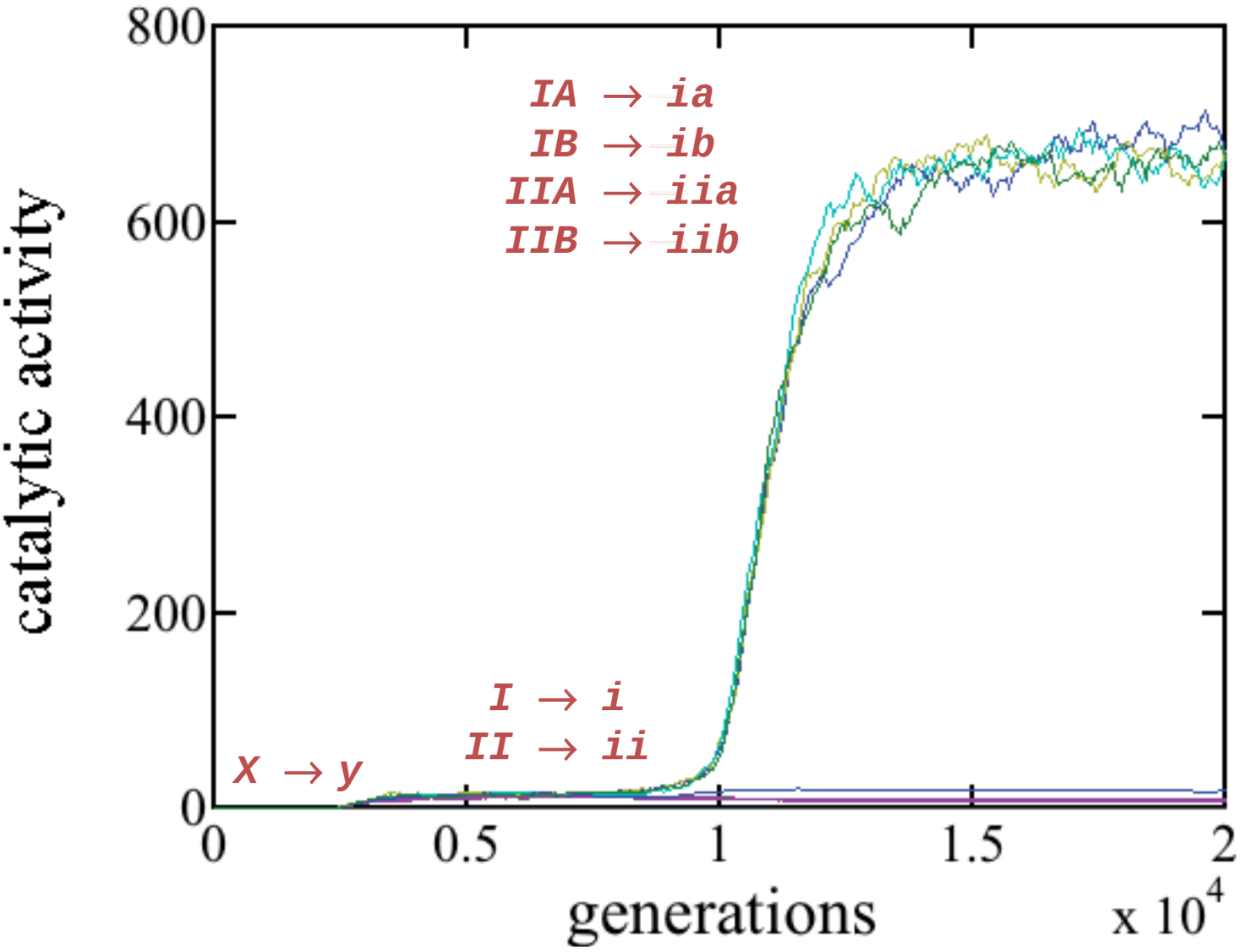
An autocatalytic mechanism of group selection.

Due to a non-equilibrium phase transition.

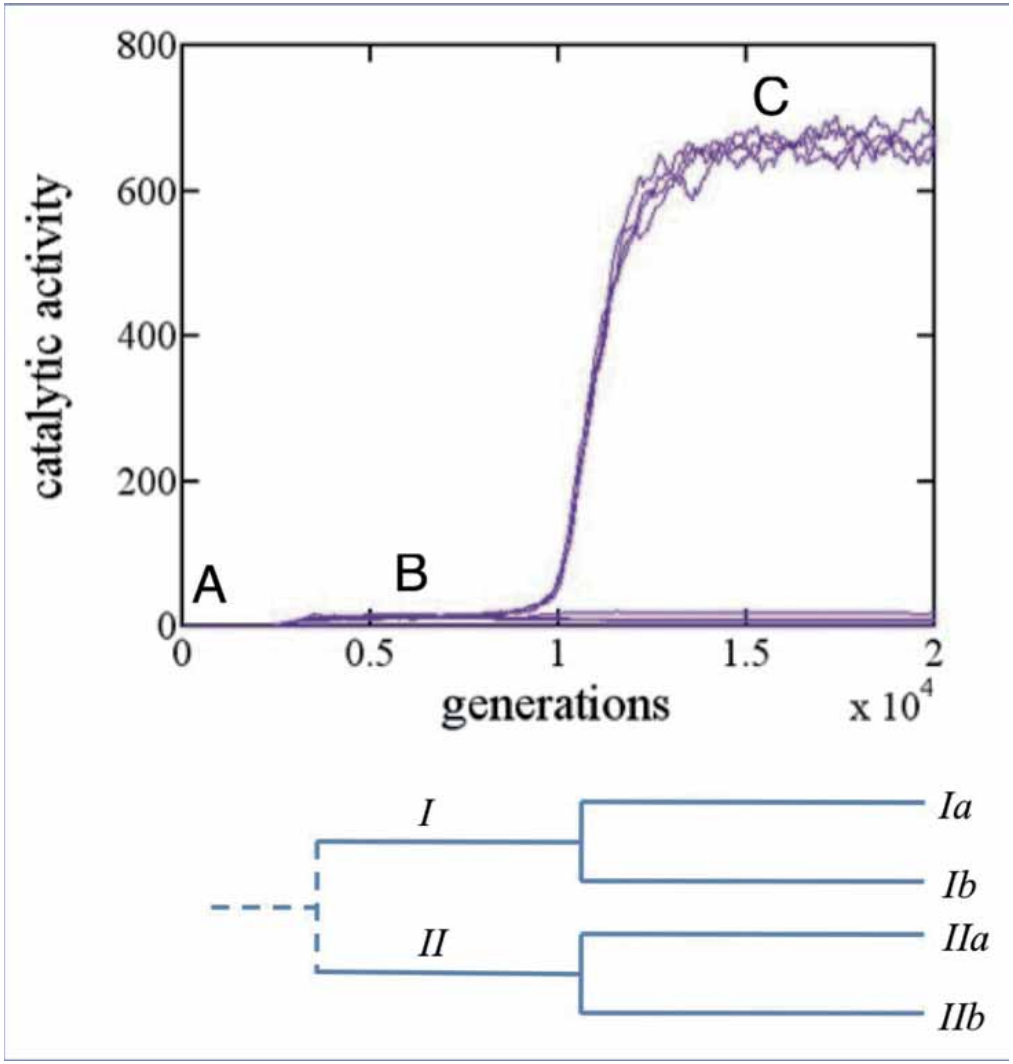

1·0
0·8
0·6
0·4
0·2
Accuracy of translation (q)
1000
800
600
400
200
0
Total catalytic activity (C_{tot})
0
500
1000
1500
2000
2500
3000
Generation number (t)

Stepwise coding self-organization

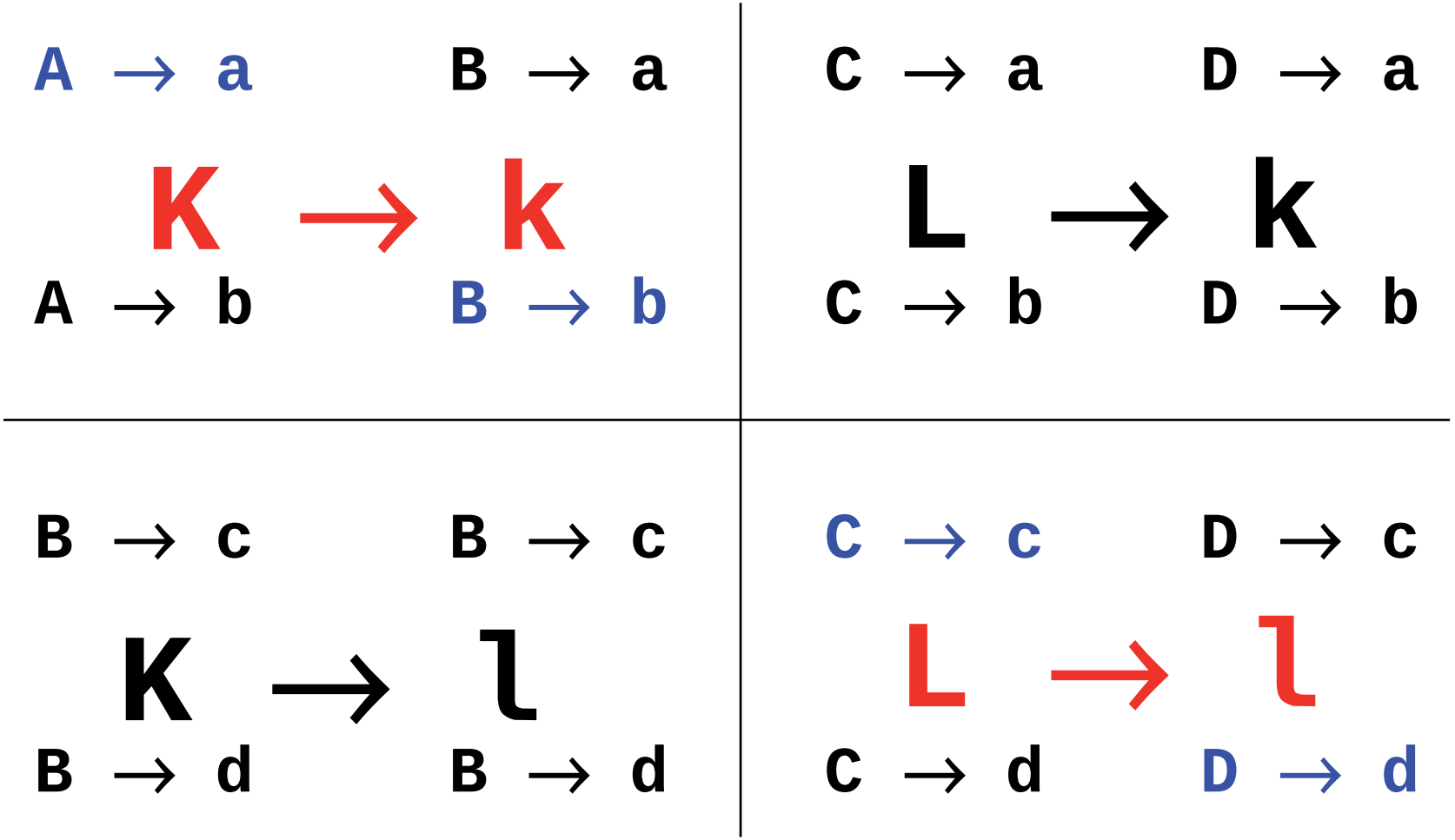
Wills P R (2004) *Stepwise evolution of molecular biological coding* in J Pollack, M Bedau et al. (eds.) **Artificial Life IX** (MIT Press, Cambridge, MA) pp51-56



Stepwise coding self-organization

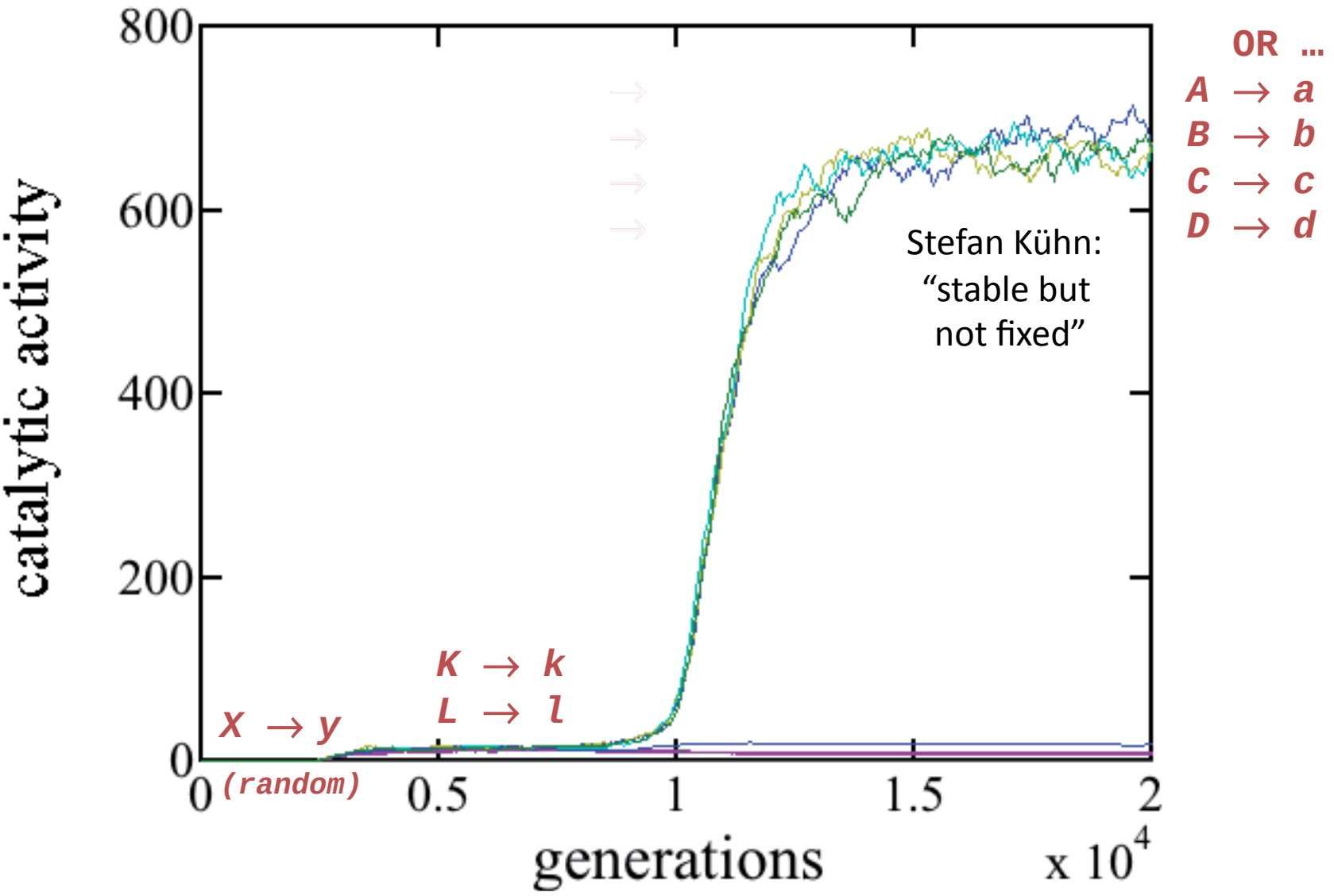


Code decomposition (2-fold)



Stepwise coding self-organization

Wills P R (2004) *Stepwise evolution of molecular biological coding* in J Pollack, M Bedau et al. (eds.) **Artificial Life IX** (MIT Press, Cambridge, MA) pp51-56



Emergence of semiosis

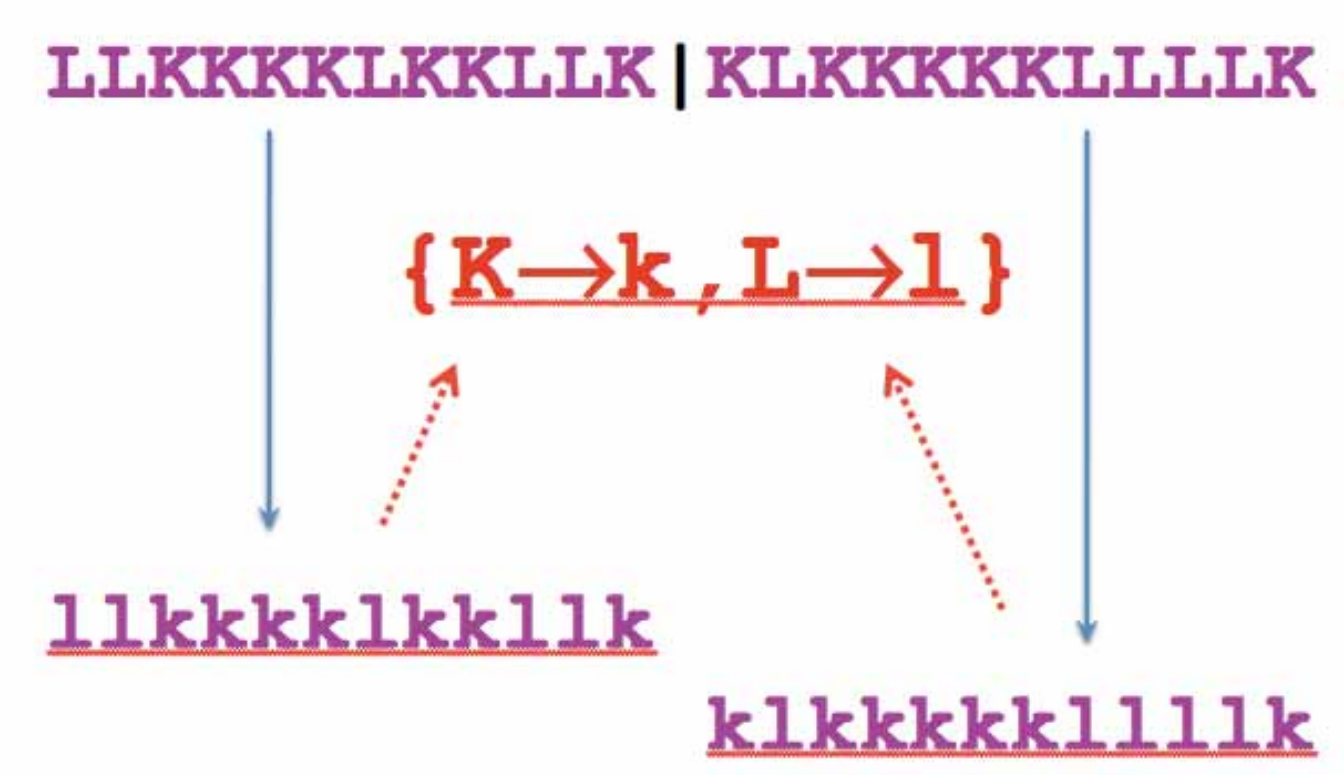
Initial state in which the sequences of the polymers being synthesised are completely random. Coding behaviour of the system is emergent, generated *de novo.*

Autocatalytic closure:

- polymers of the sort being synthesised play a role in the synthetic process that produces them
- polymers serve as adaptors,
  influence choice of monomer to be concatenated
  recognition of monomer present at the collinear position on the template.

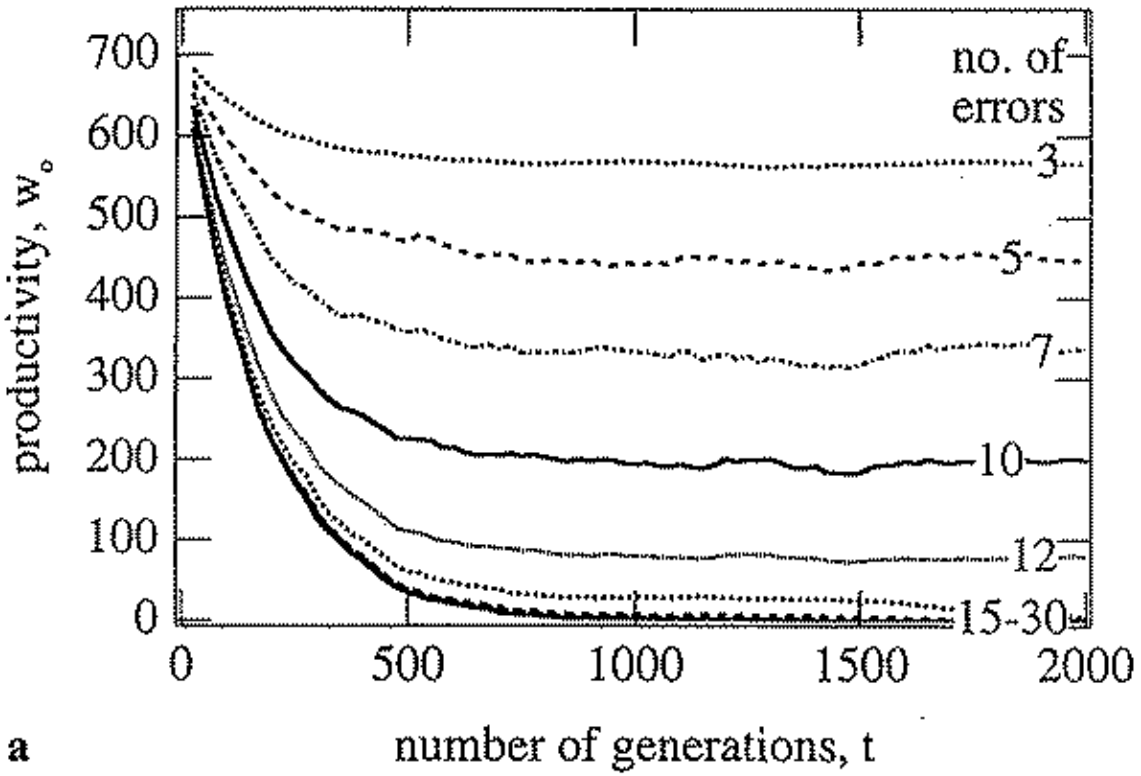
What is wrong?

We are relying on the “miracle” of having the correct reflexive information



What is wrong?

The information would slowly decay, and with it the required reflexivity



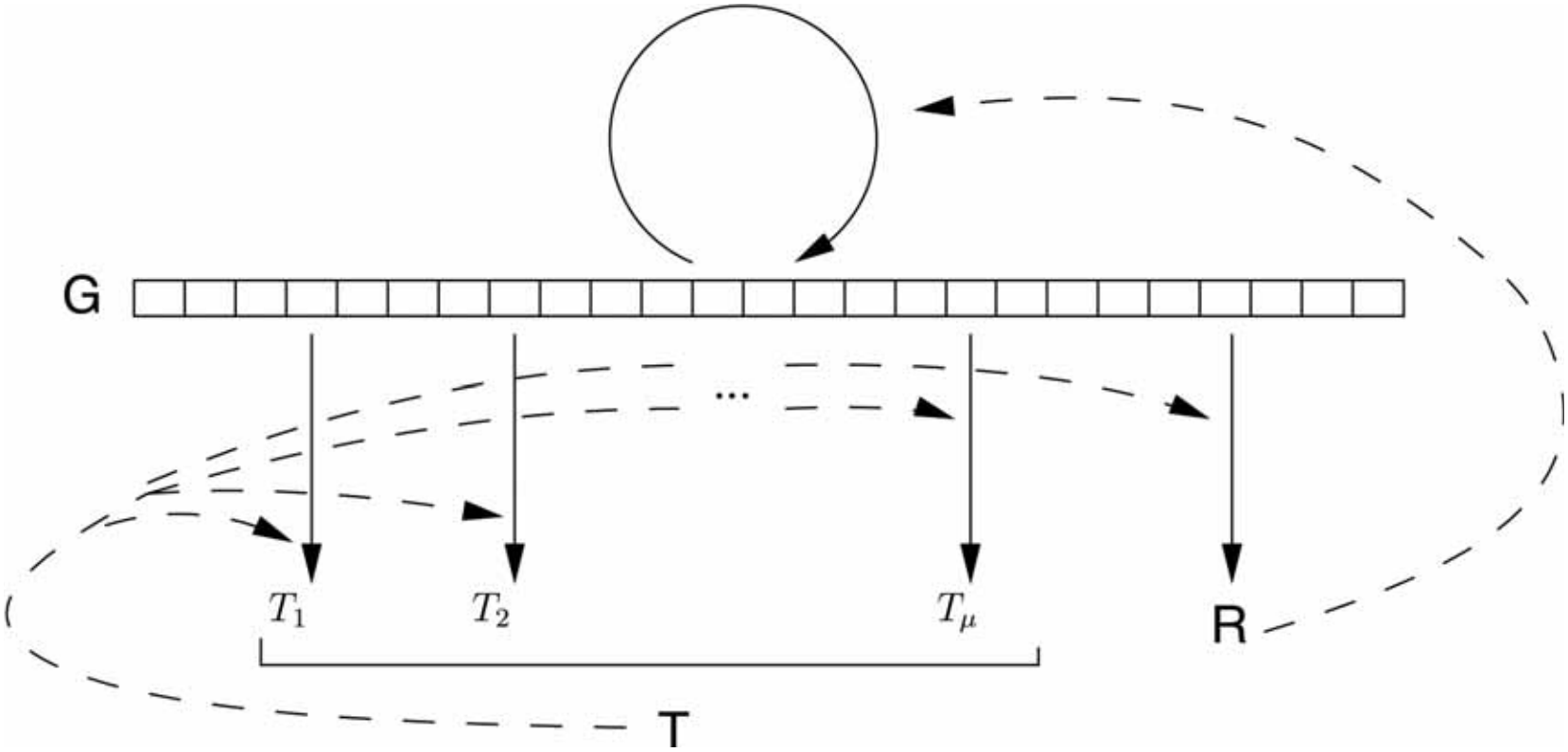
What is wrong?

The information would slowly decay, and with it the required reflexivity,

so how might we explain that organisms actually exist?

Let’s look at model “GRT systems”.

Gene-Replicase-Translatase systems



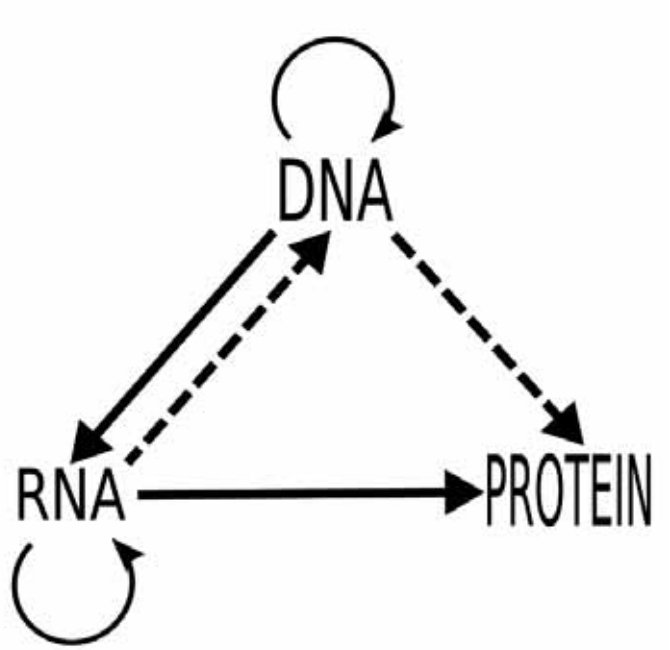
GRT systems (Füchslin & McCaskill)

What determines biological specificity?

Crick (1957; 1970)

sequence hypothesis

Central Dogma



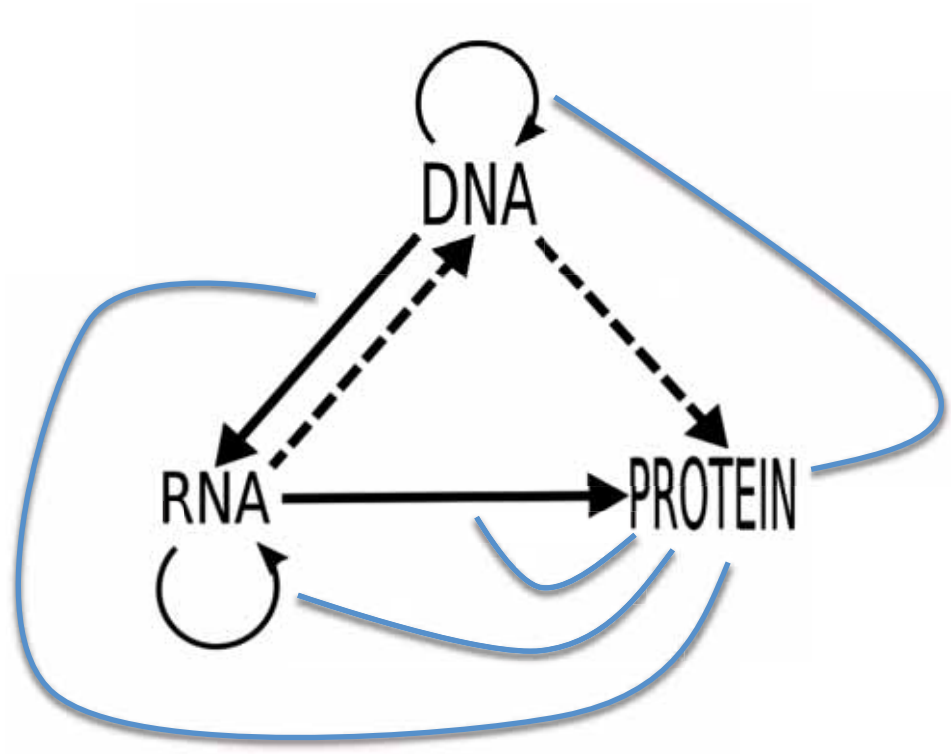
What determines biological specificity?

Crick (1958; 1970)

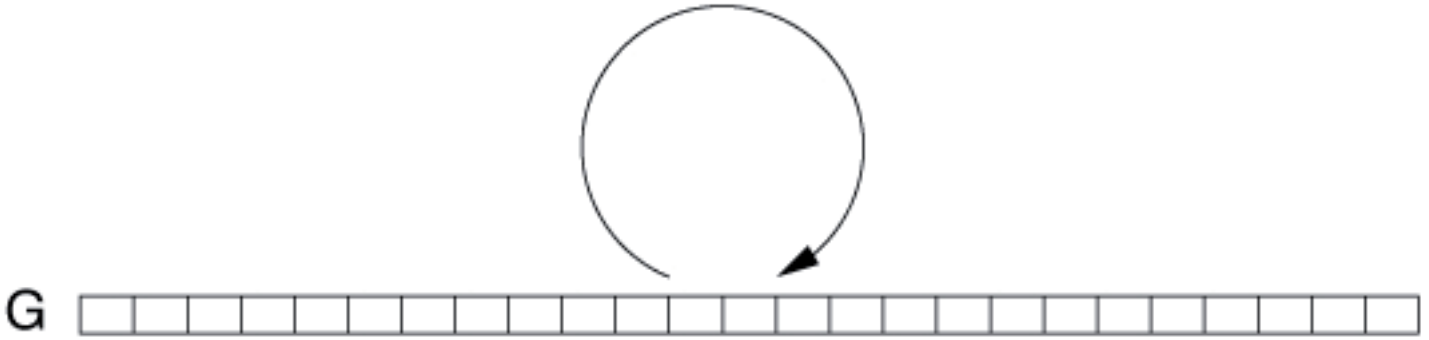
sequence hypothesis

Central Dogma

Represents catalytic function of proteins

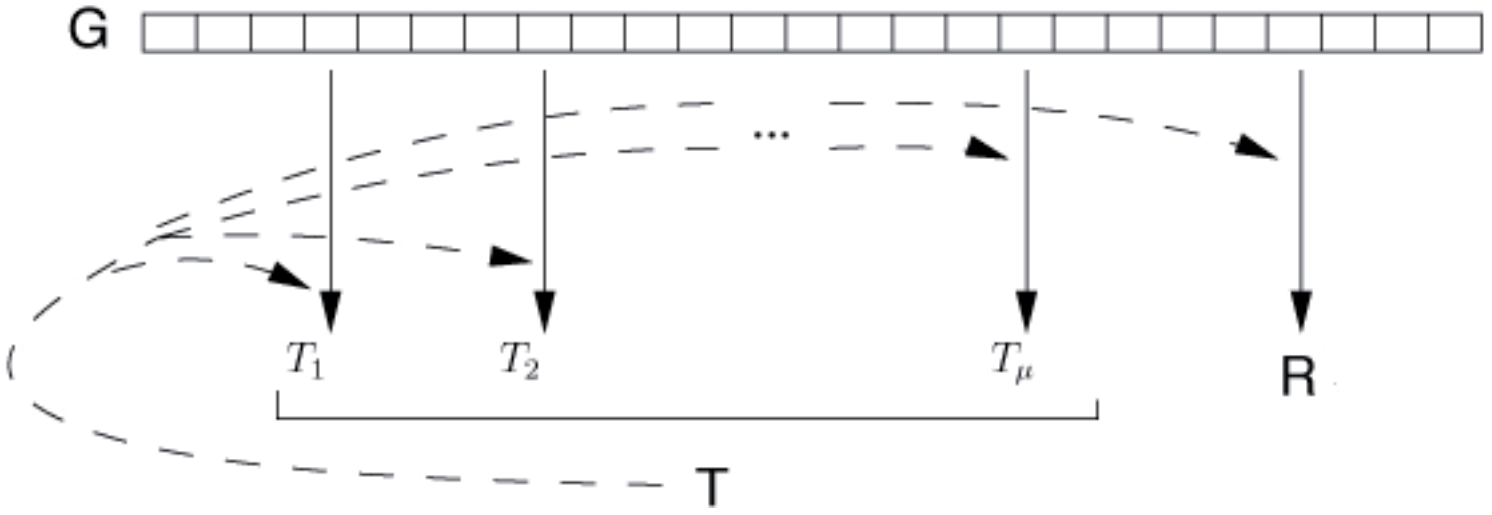


GRT systems
Replication of Genes



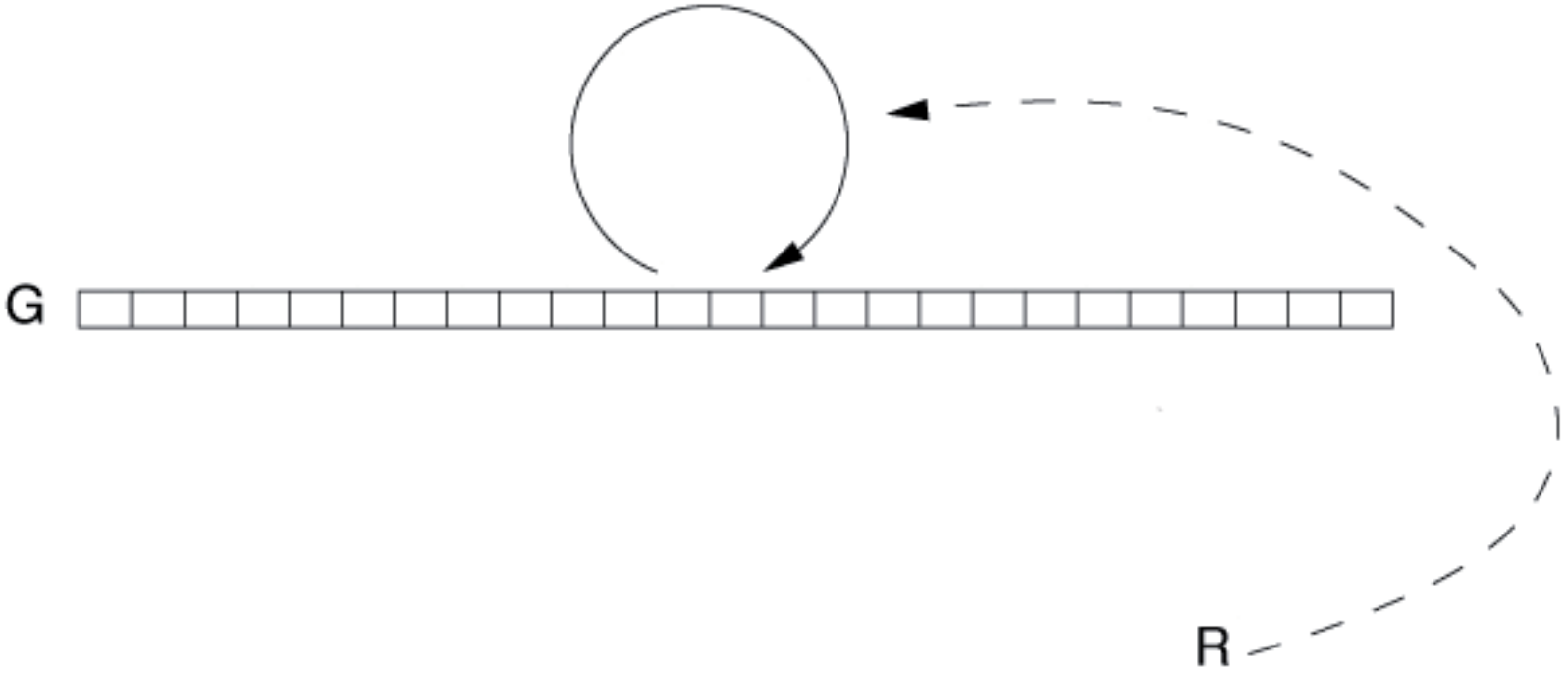
Darwinian selection of information carriers

GRT systems
Synthesis of Translatases



Self-organisation of coding autocatalysis

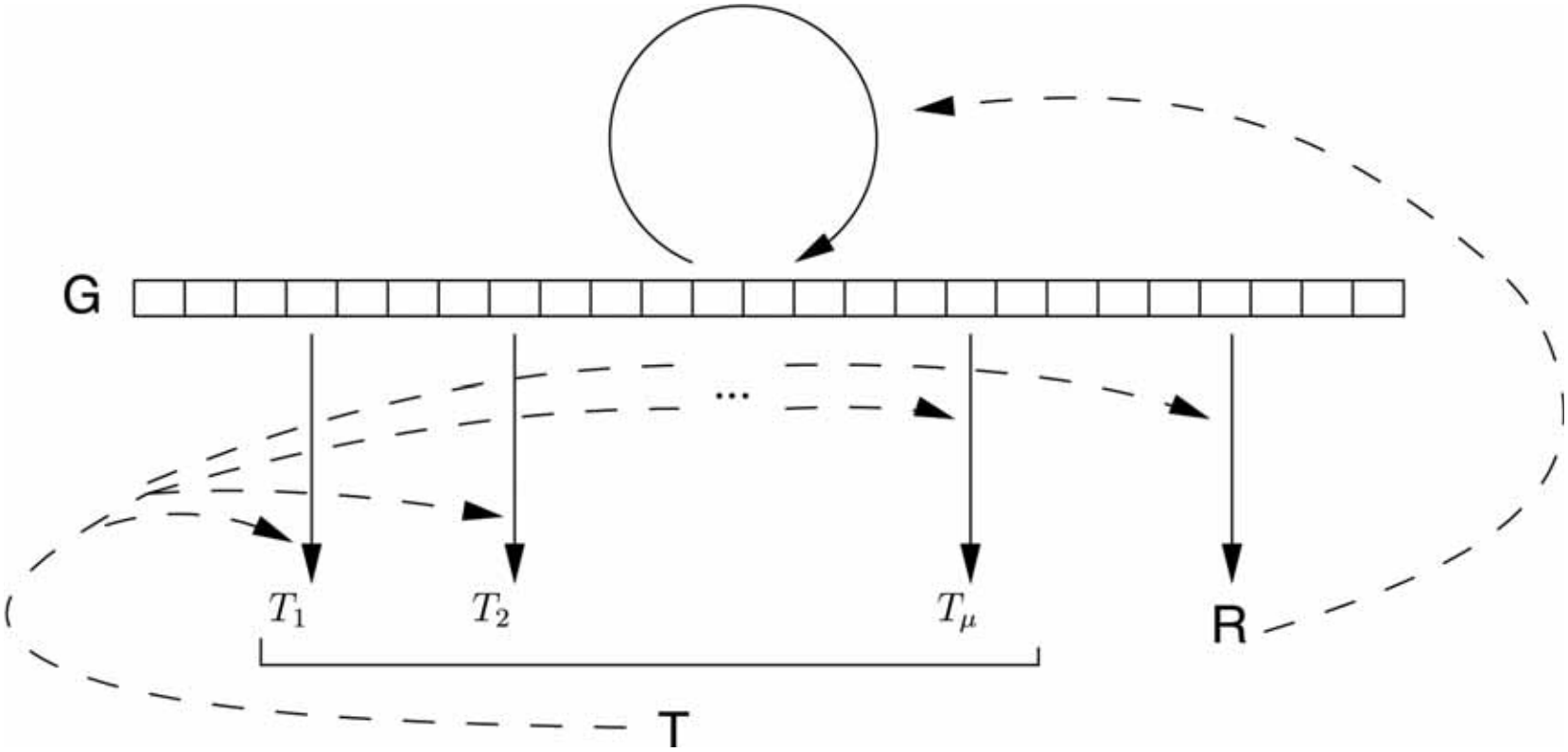
GRT systems
Replicase is systemic



Functional autonomy

GRT systems

Homogeneous GRT system is impossible.



Note: all the networks discussed today have been modeled as homogeneous systems.

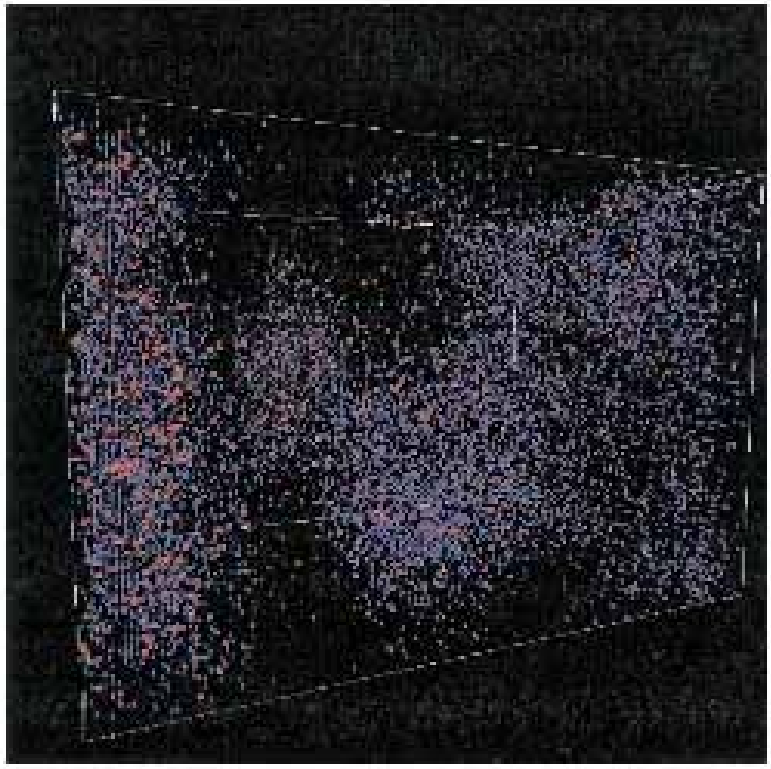
Coding systems need spatial stabilization

Füchslin, R.M., McCaskill, J.S., 2001. Evolutionary self-organization of cell-free genetic coding. Proc. Natl. Acad. Sci. USA 98, 9185–9190.

Coding systems:

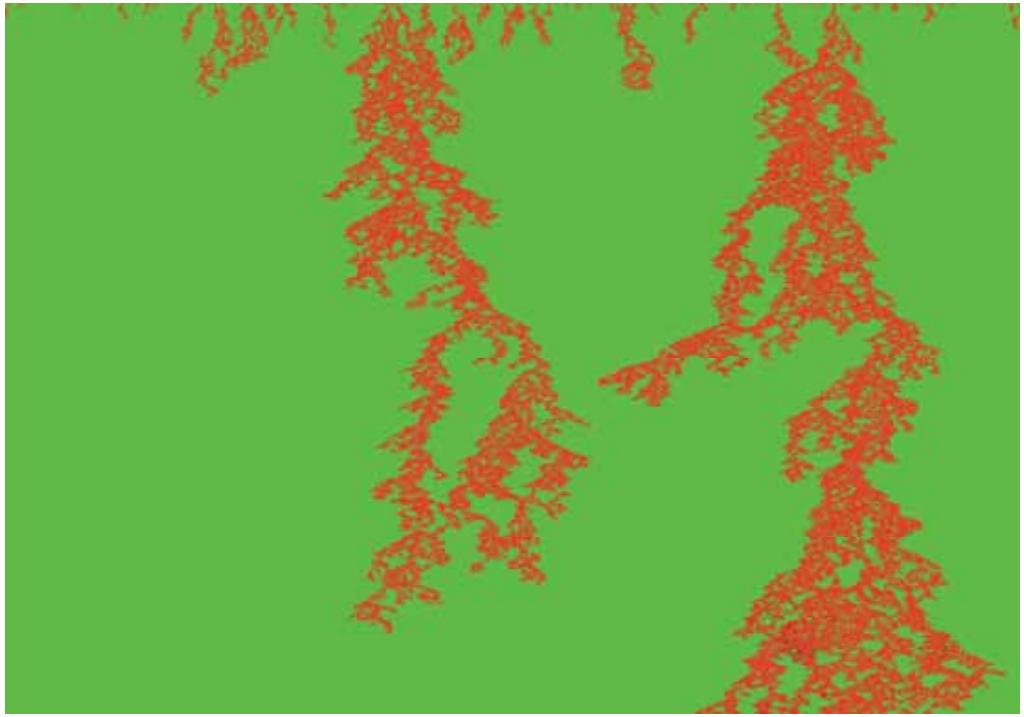
- unstable in homogeneous solution

- stabilization through reaction-diffusion coupling (Turing mechanism)



GRT self-organization in a 1D compartmentalized system

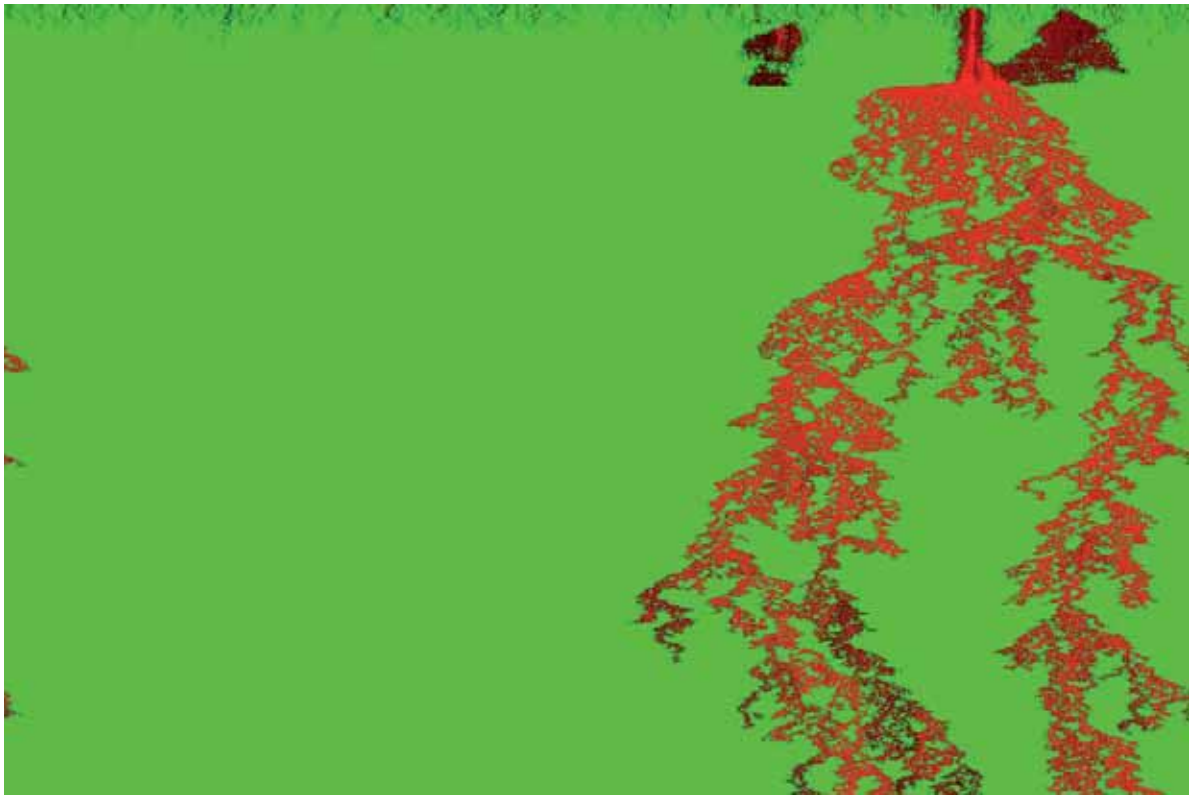
Markowitz S, Drummond A, Nieselt K & Wills P R (2006) *Simulation model of prebiotic evolution of genetic coding*, in L M Rocha et al. (eds.) **Artificial Life X** (MIT Press, Cambridge, MA) pp152-157



GRT Turing-type dissipative structure

space →

time ↓



GRT systems
Replicase is systemic

Theses systems are quite different from other model prebiotic systems in the way that they use information – all the information molecules are energetically and functionally equivalent.

Thank you!

p.wills@auckland.ac.nz