Genotype-Phenotype Maps:
from molecules to simple cells

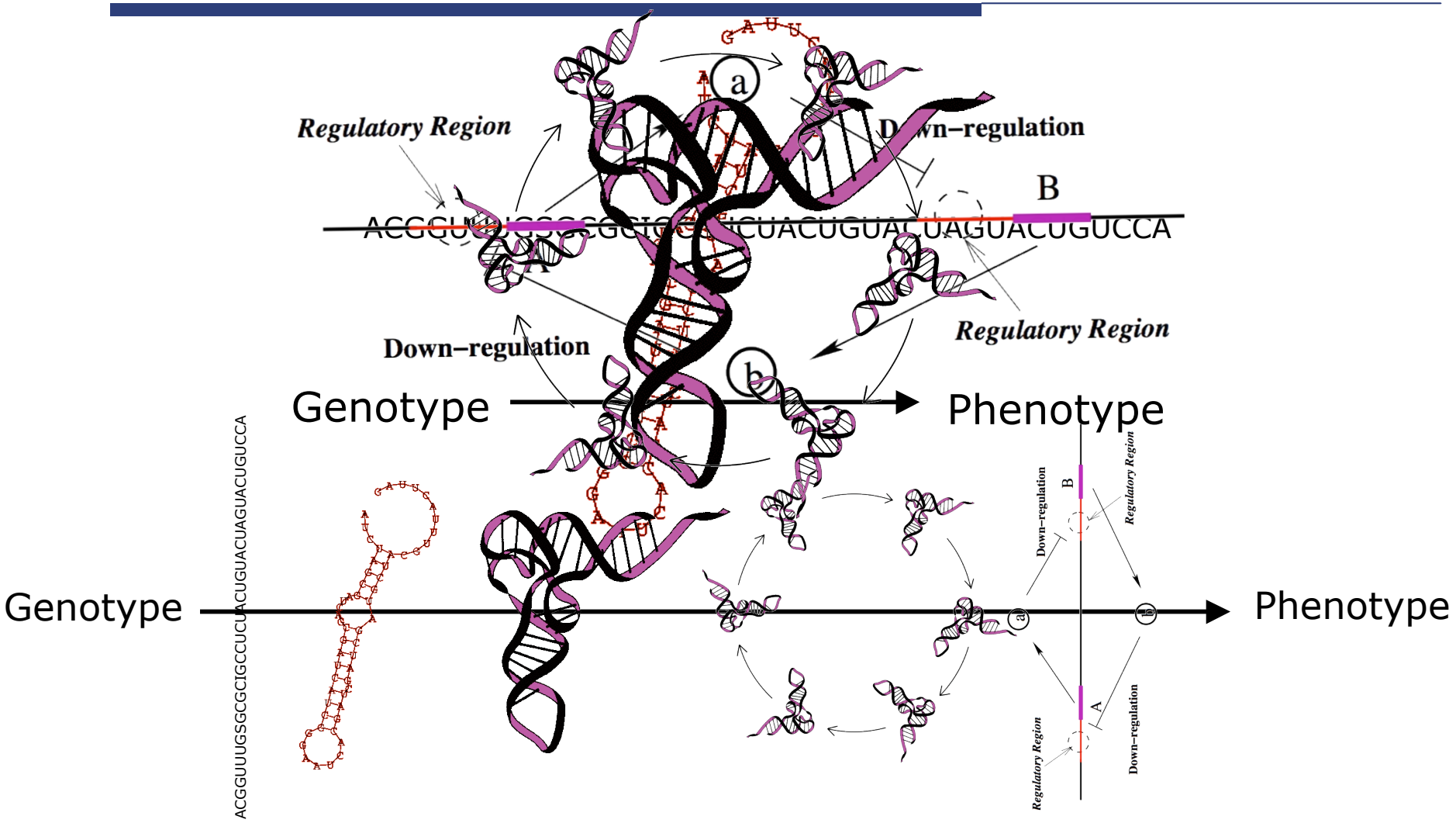
Camille Stephan-Otto Attolini

Institute for Theoretical Chemistry,

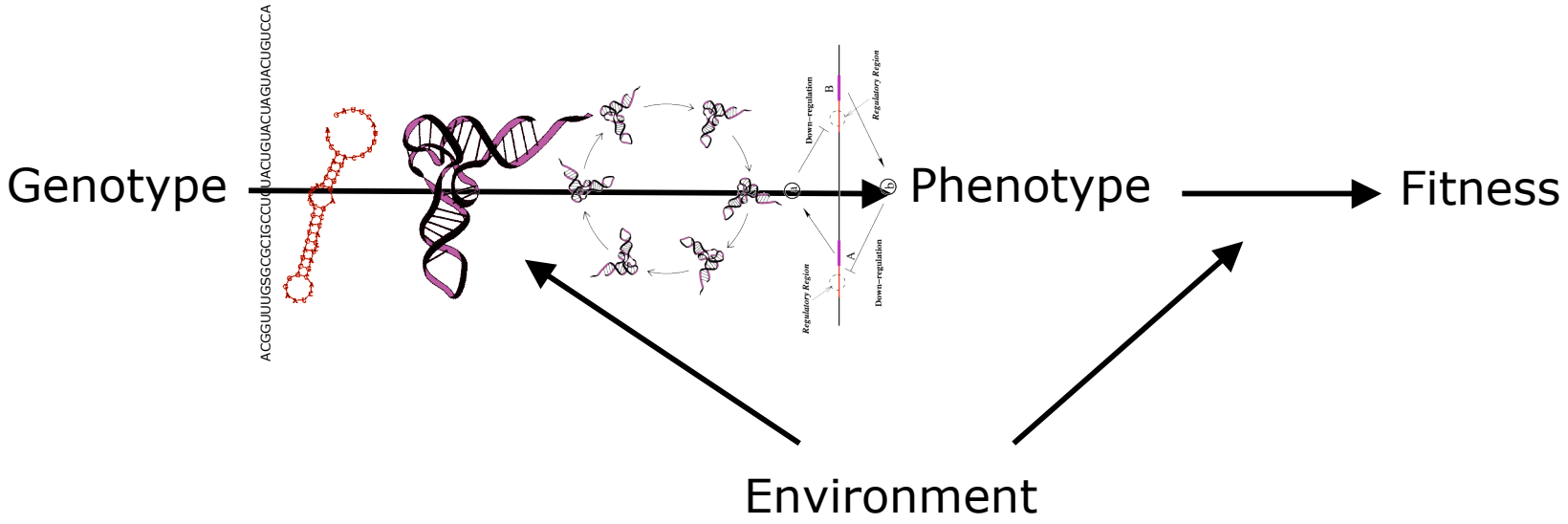
University of Vienna, Austria

Vienna, February 28th, 2006

The Genotype-Phenotype Map

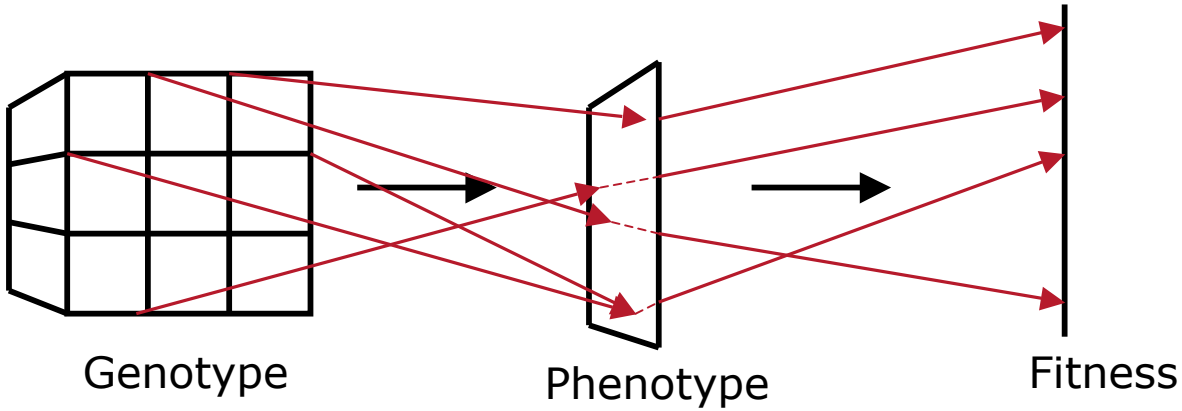


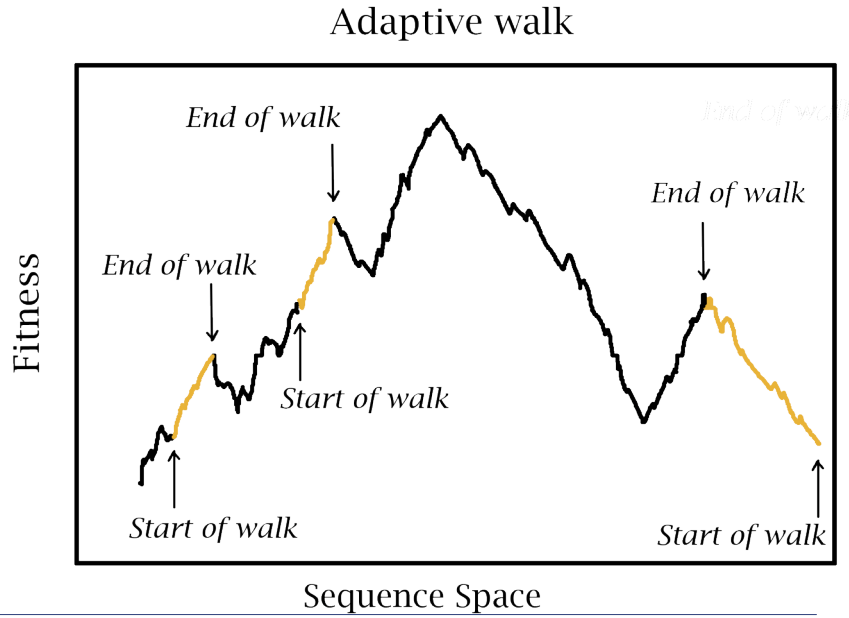
The Genotype-Phenotype Map II



$$G \times E \xrightarrow{f} P \times E \xrightarrow{g} F$$

Genotype, Phenotype and Fitness





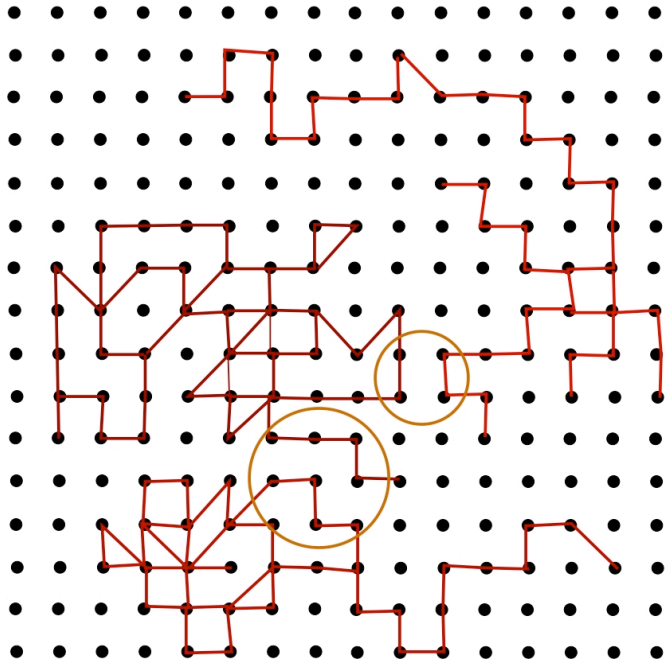
Characteristics I

Neutrality

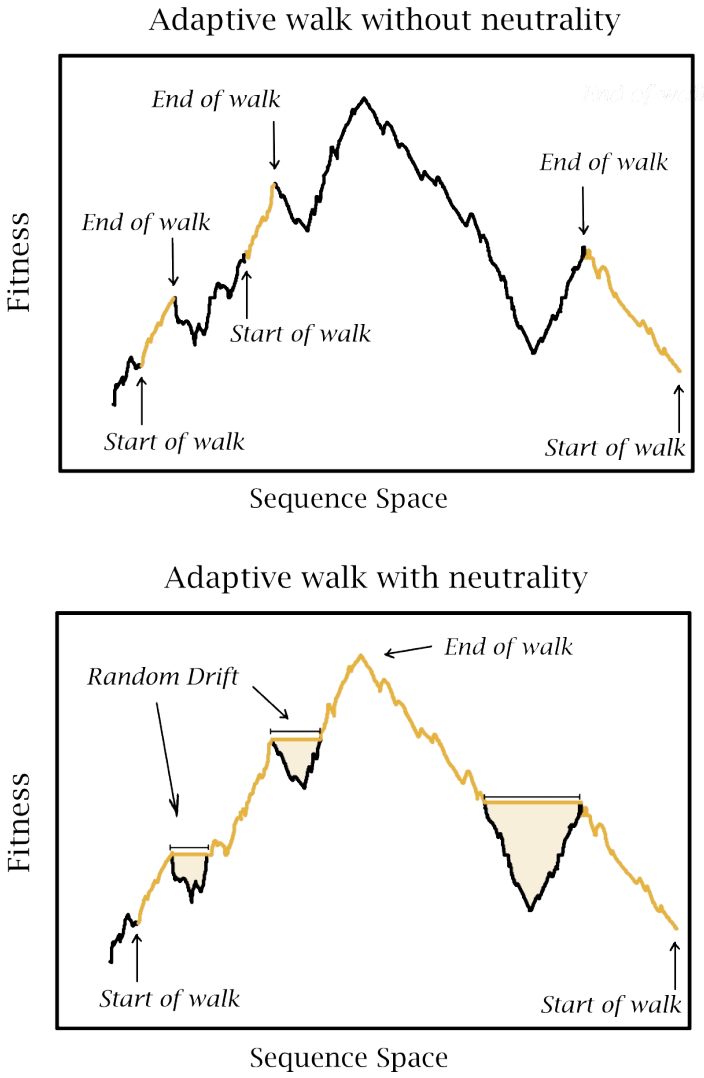
ACGGUAGCGAUCGA**C**AUCAGCGAUCAGC → ACGGUAGCGAUCGA**U**AUCAGCGAUCAGC

Characteristics I (bis)

Neutrality and the fitness landscape



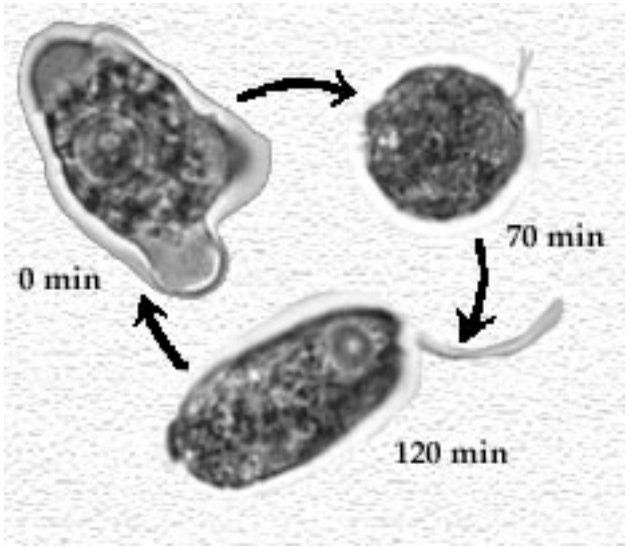
Neutral networks



Characteristics II

Plasticity

- Is the capability of a phenotype to change due to impulses of the external conditions. In higher order organisms it is controlled by the genotype.



Amoeba Naegleria Gruberi

Characteristics III

Variability and Evolvability

- Variability is the capability of a genome to change, usually due to errors in the copying process. Without this, evolution would be impossible.

- An evolvable genotype has the possibility of finding fitter phenotypes.

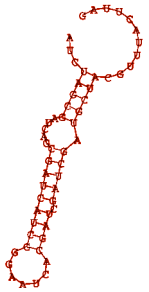
It has been shown by P. Schuster and coworkers that the RNA sequence to secondary structure map has all these characteristics…

Levels

Molecular Level

ACUUGSGCGCIGCCUCUACUACUAGUACUGUCCA

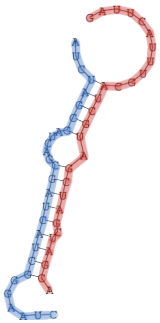
RNA secondary structure

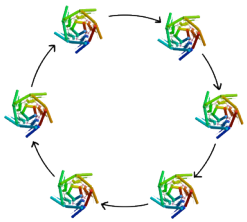


Multi-Molecular Level

ACUUGSGCGCIGCCUCUACUACUAGUACUGUCCA

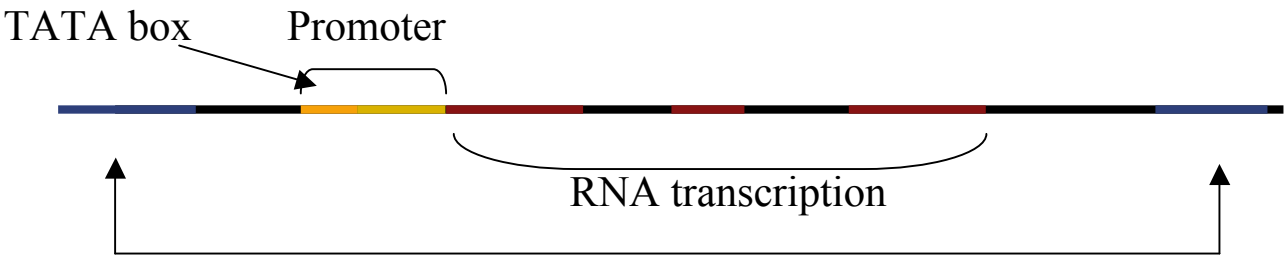
RNA cofolding





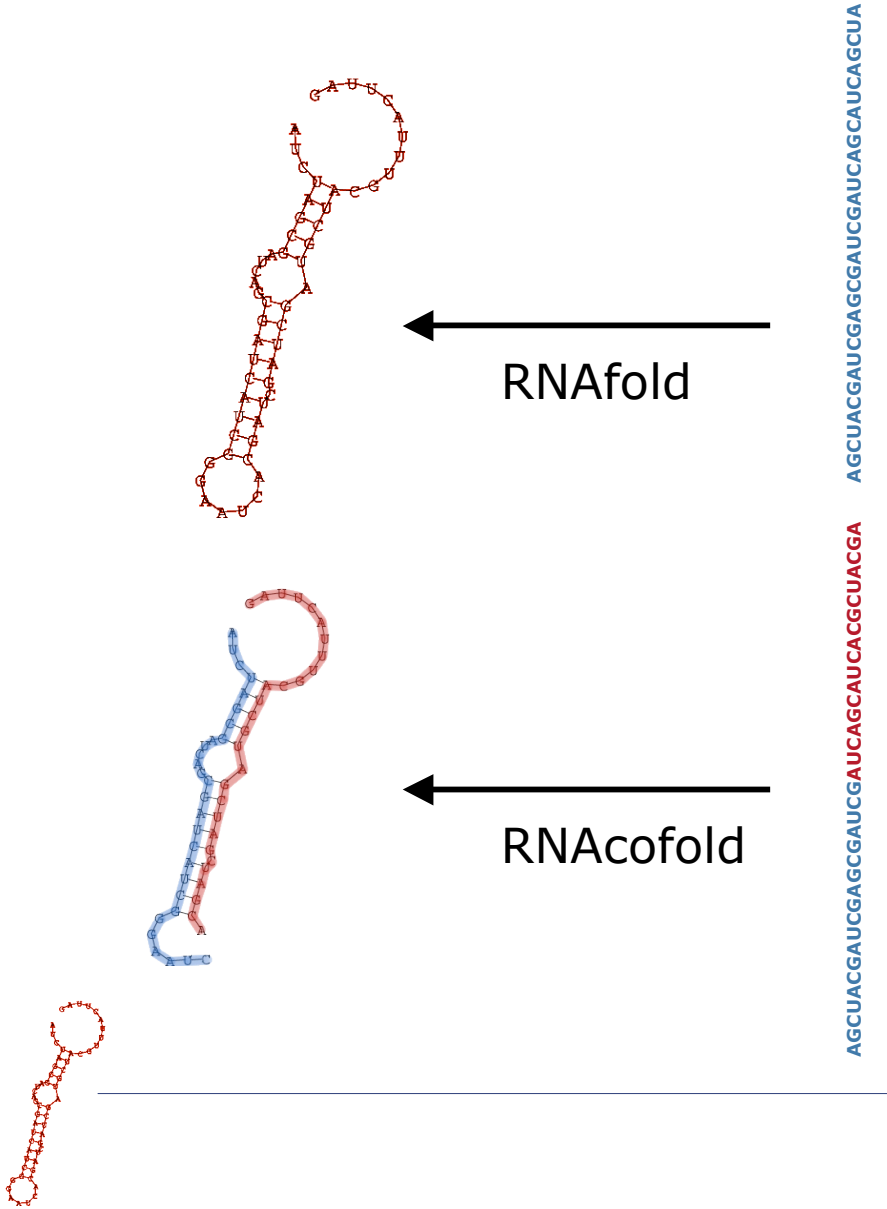
Hypercycle

Regulation



Regulatory regions

At the molecular level

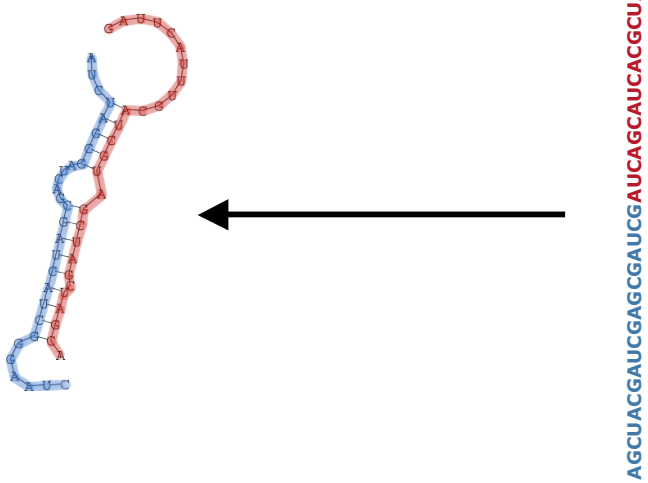


The secondary structure of a given RNA sequence is the list of (Watson-Crick and wobble) base pairs such that:

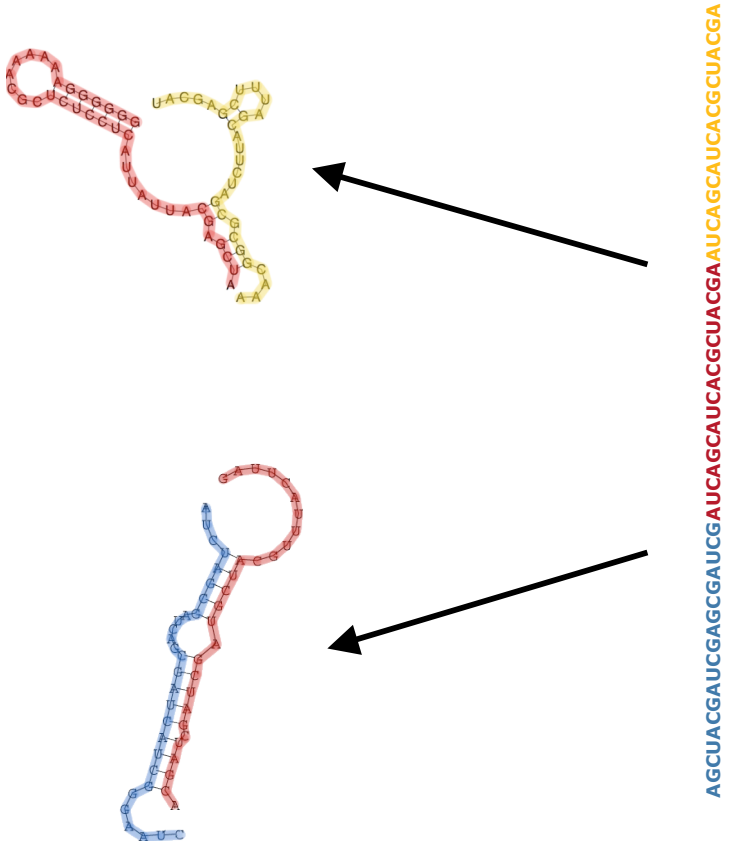
- each nucleotide takes part in at most one base pair, and
- base pairs do not cross (knots and pseudo-knots are forbidden)

The RNAcofold algorithm computes the hybridization complex of two RNA molecules. The energy of the loop where the two sequences meet is computed as an external loop.

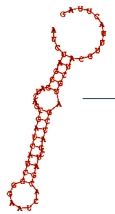
Neutrality in RNAcofold



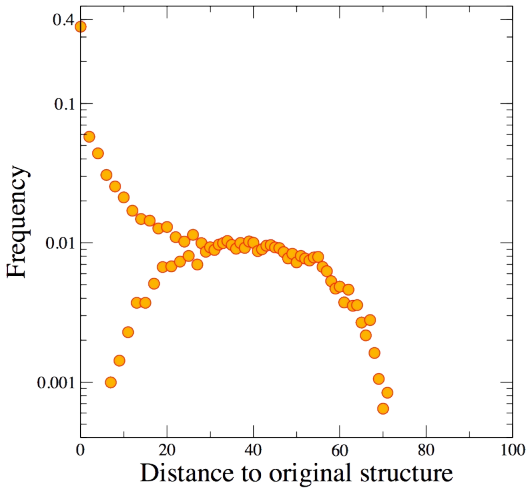
Two sequences



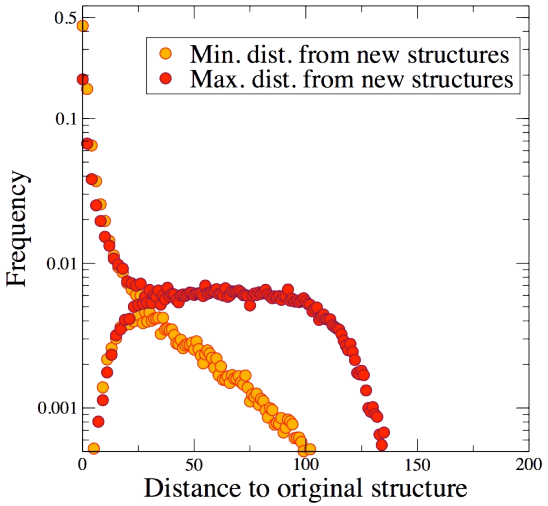
Three sequences



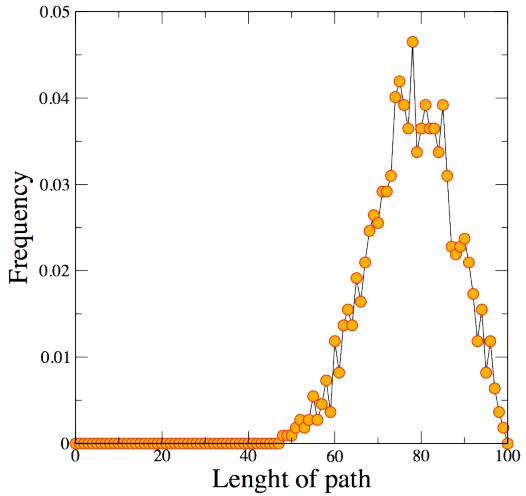
Neutrality and length of **neutral paths** in RNAcofold

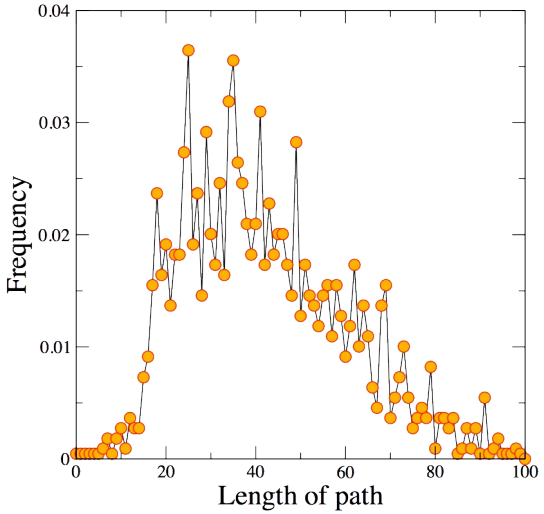


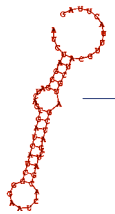
Frequency of neutral mutations:
0.32



Frequency of neutral mutations:
0.18





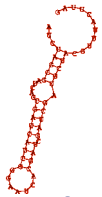
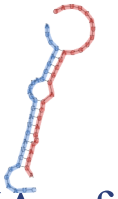


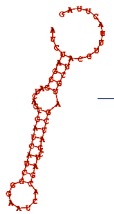
Two sequences

Three sequences

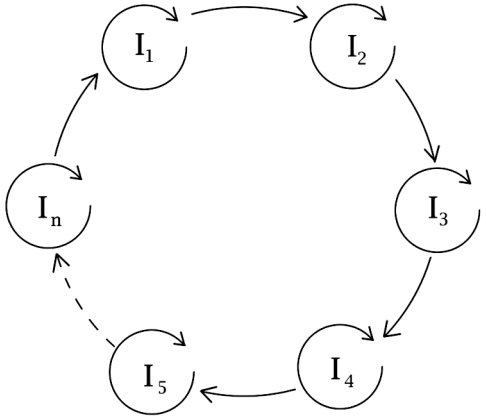
Length of sequences:
100 bases

RNAfold vs. RNAcofold

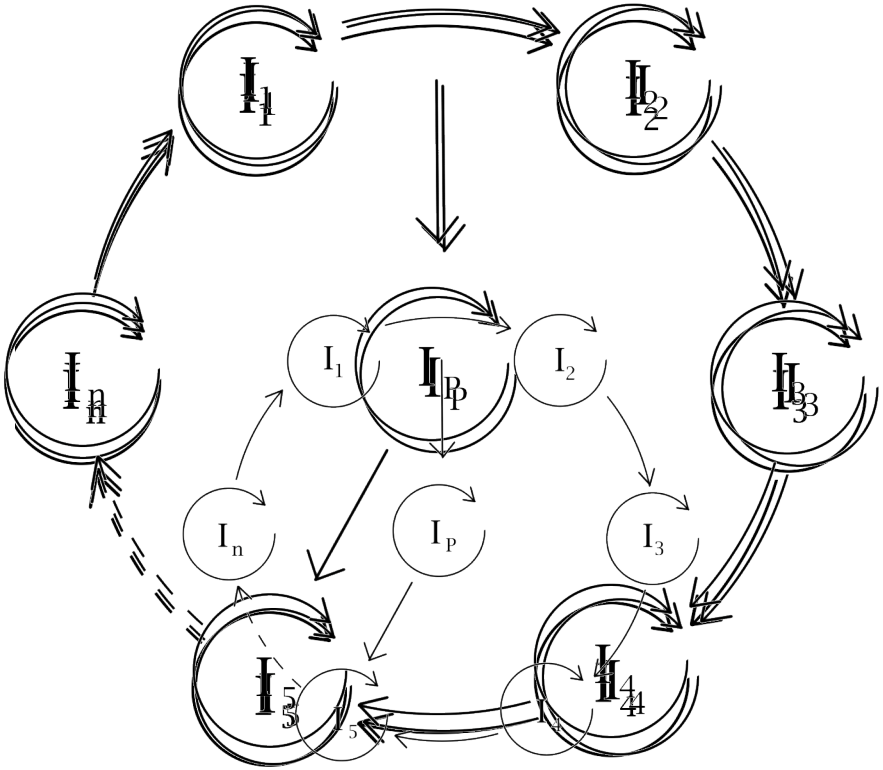
	Sequences	Neutrality	Length of N.P.
 RNAfold	One sequence	0.33	100
 RNAcofold	Two sequences	0.32	75
RNAcofold with two different sequences	Three sequences	0.18	40



Multi-molecular level

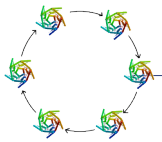


The Hypercycle



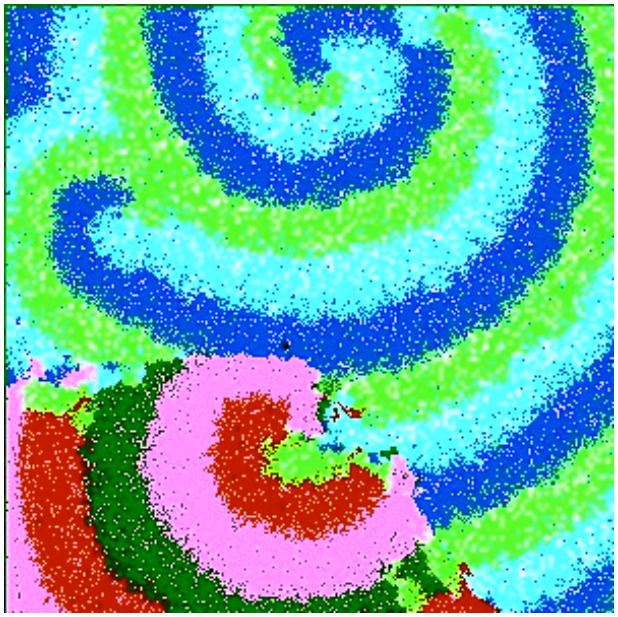
Selfish parasite

The copying process of molecules as the RNA allows only a few dozens of nucleotides in length, due to the high rate of errors. The hypercycle was proposed by Eigen and Schuster in '79 as an answer to the Selfish parasite Catch-22.

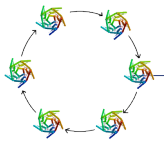


The Hypercycle in two dimensions

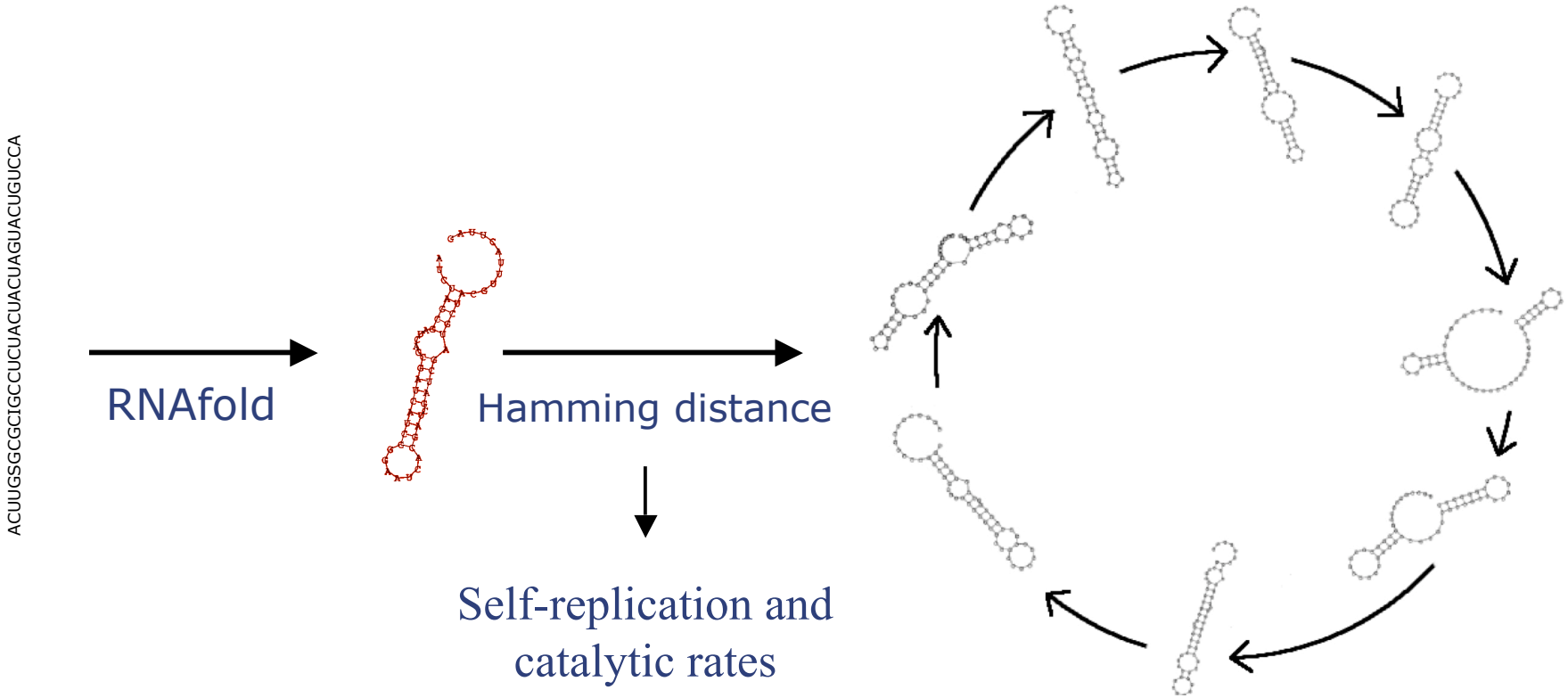
Hogeweg and Boerljist proposed in the late 80's a model of the hypercycle in a two dimensional lattice which resulted to be resistant against selfish parasites.

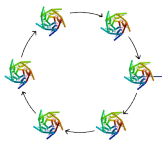


The main characteristic of this model, was the formation of spirals. This spatial organization is responsible of the expulsion of the parasite outside of the system.

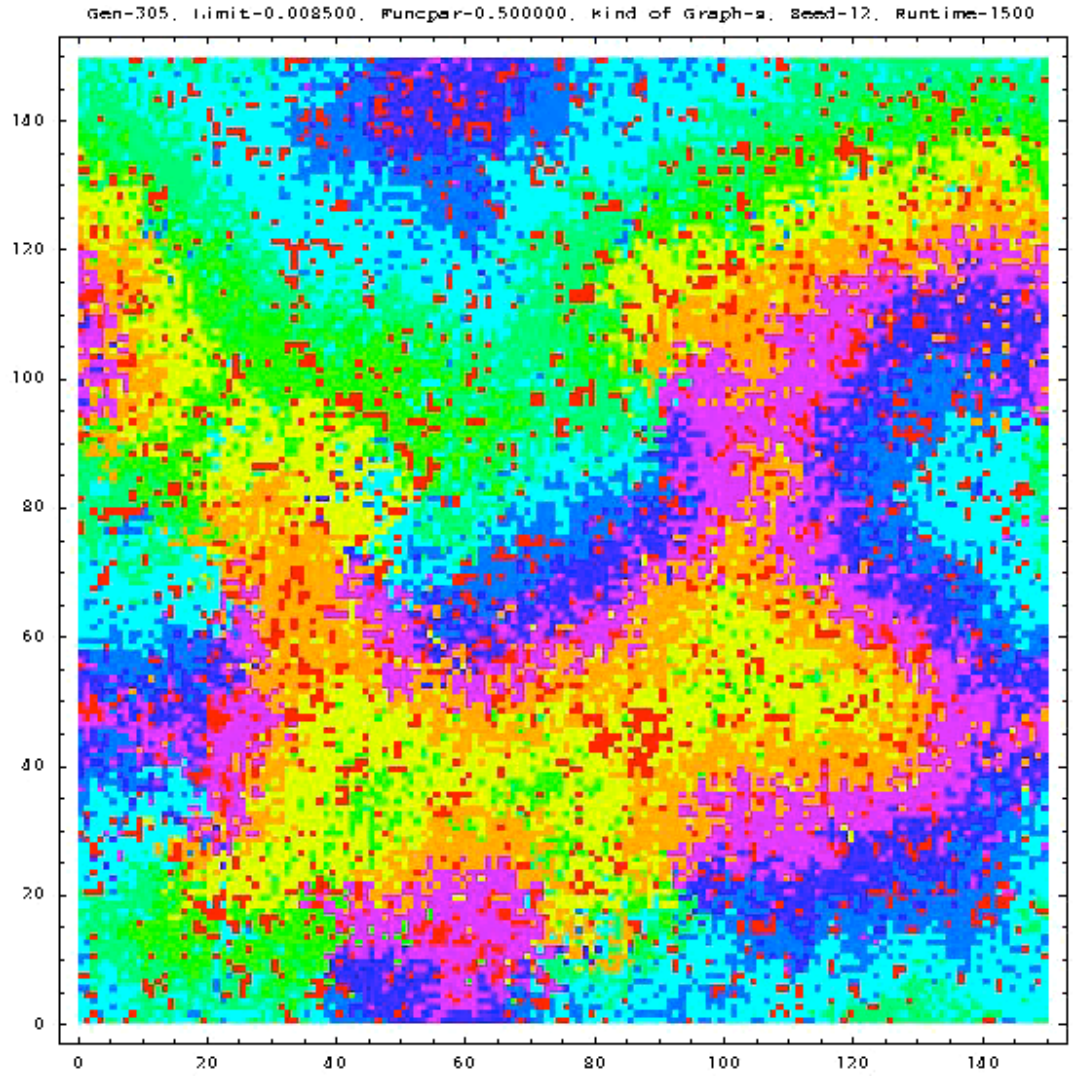


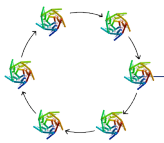
Evolving towards the hypercycle



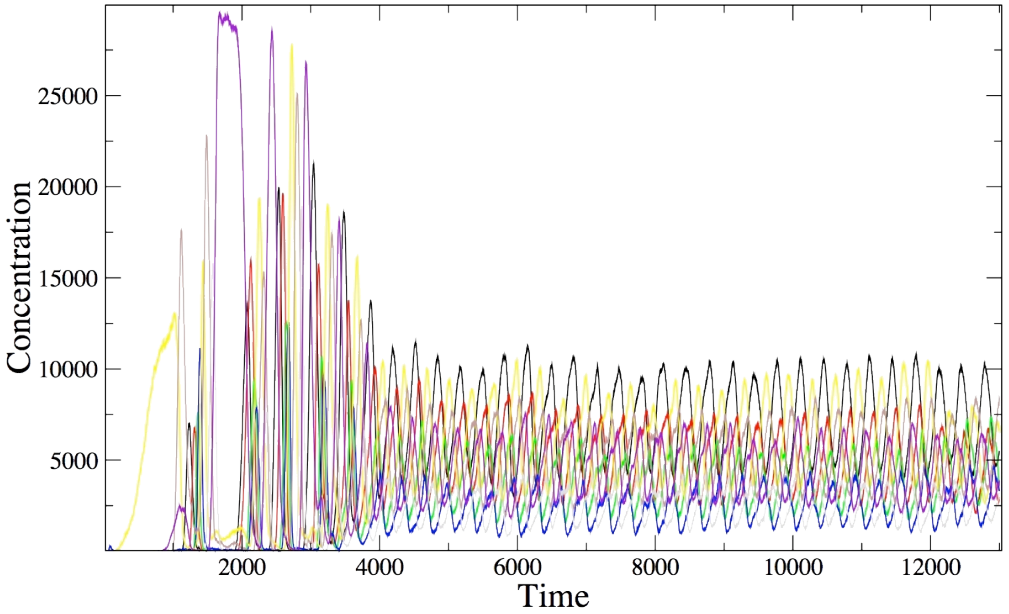


Hypercycle in two dimensions



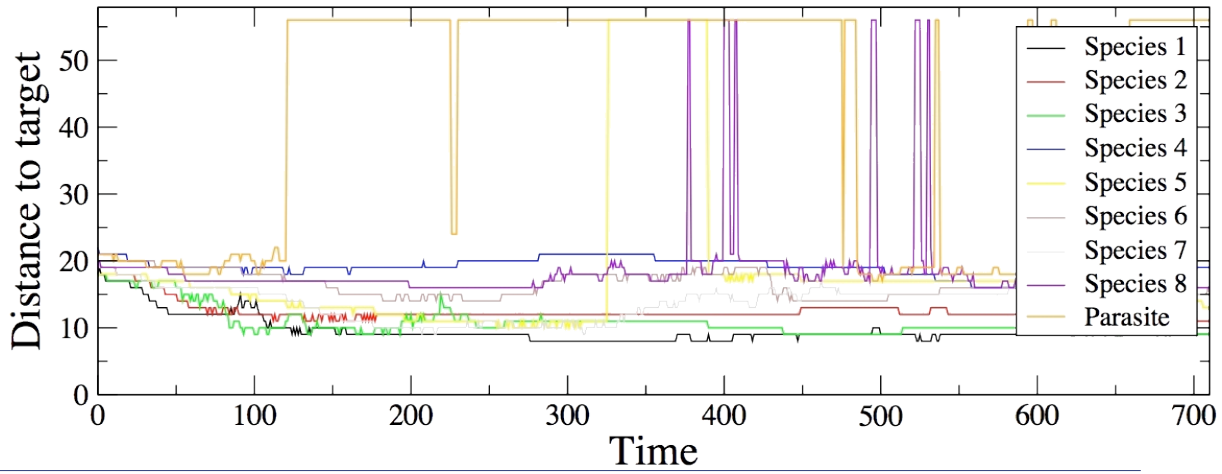


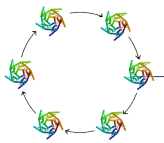
Concentration and Fitness



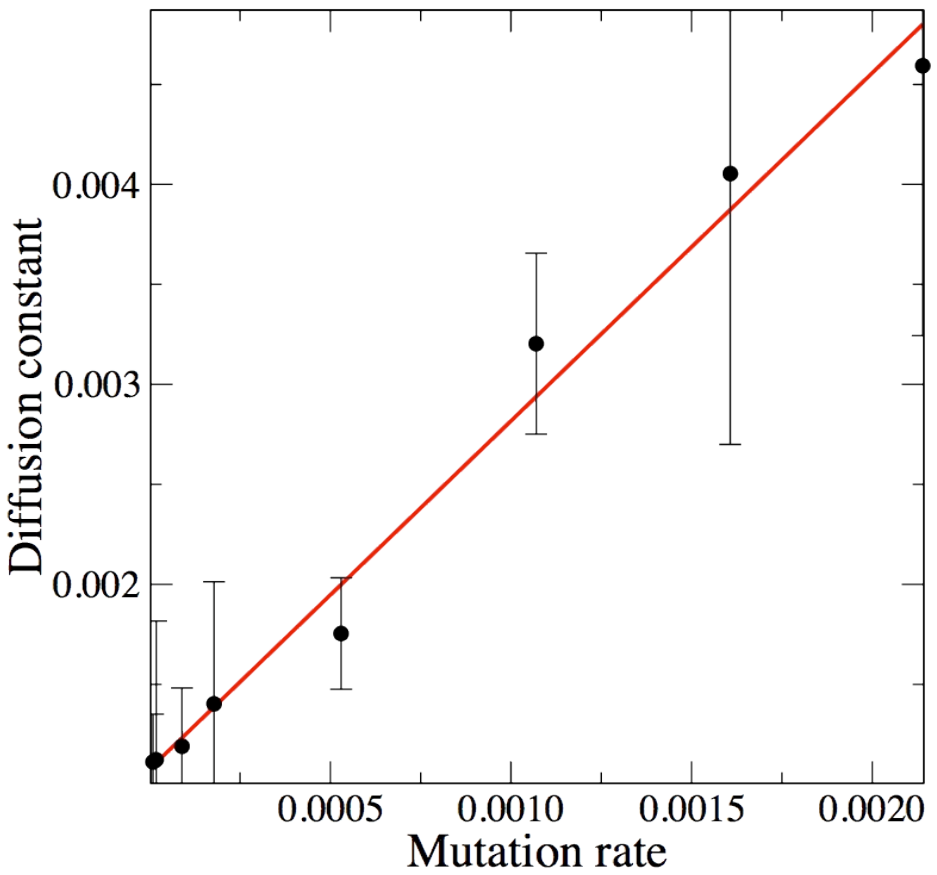
Concentration of molecules corresponding to different targets

Fitness evolution

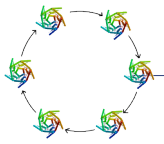




Diffusion in sequence space



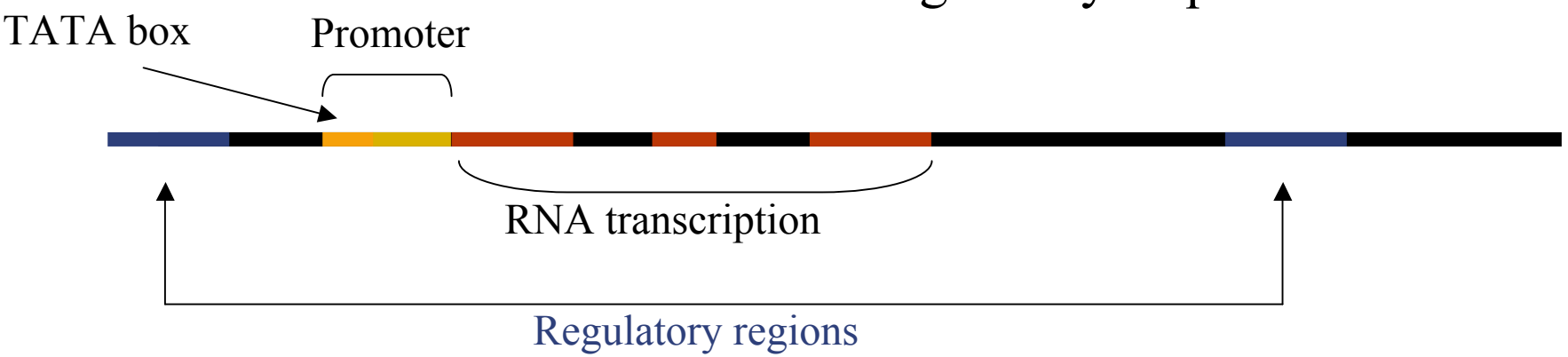
Diffusion in sequence space



Gene regulation

The regulation of a gene expression depends on multiple factors. A typical Eukaryote gene has the following structure:

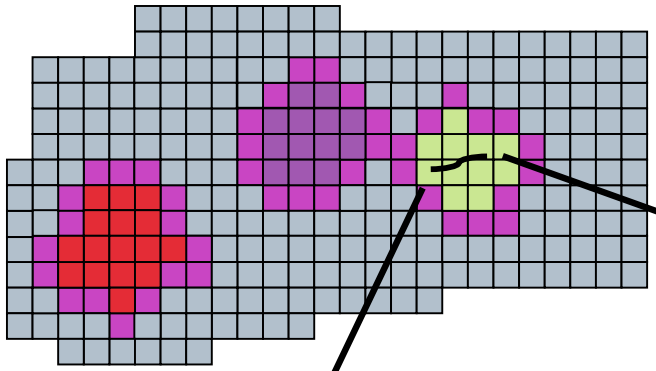
- Coding region
- Promoter
- Regulatory sequences

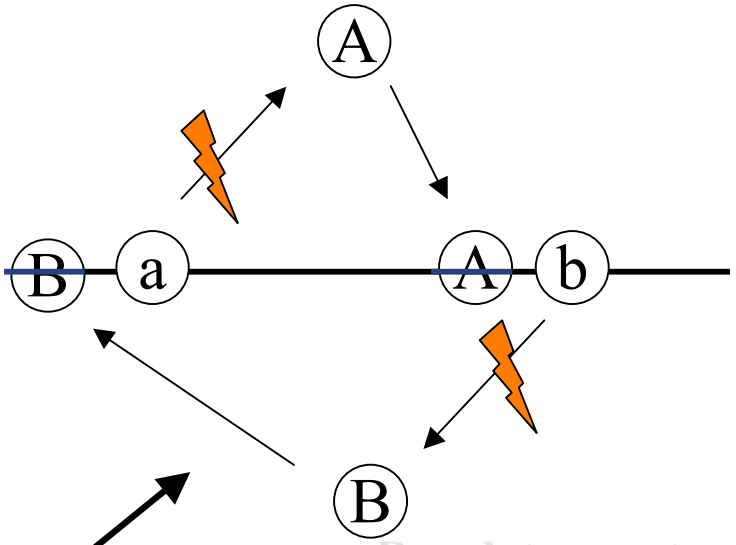


The CelloS model

Phenotype and environment

- Spatial environment with changing food sources
- Finite life time for cells and reproduction by division





Regulatory network

- Mutually repressing genes
- Cost for expressing proteins

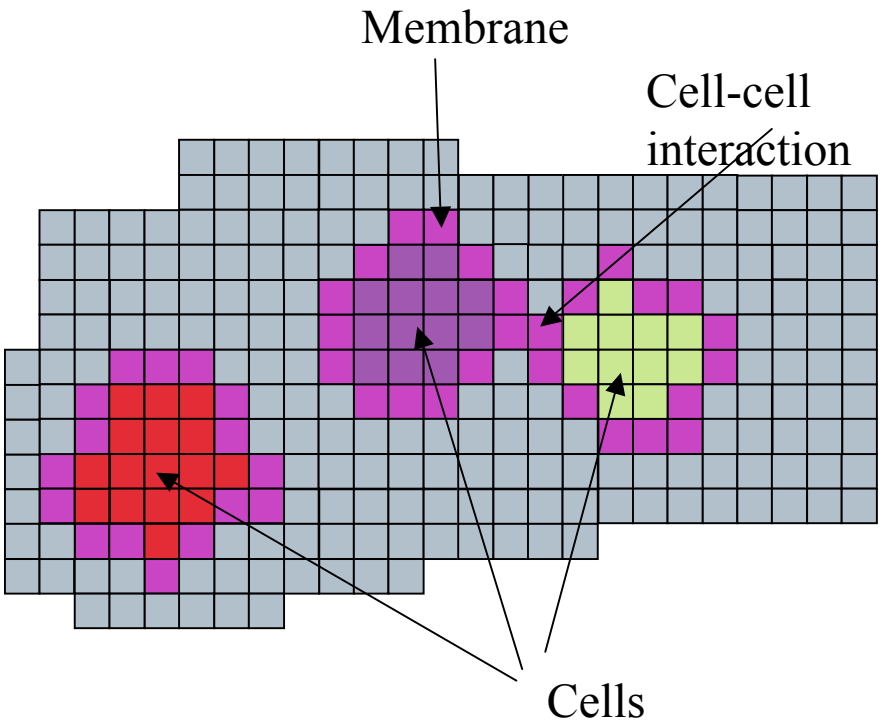
UAGCAUCAGGCGAUCAGCGAUCAUCGAUCAUCACGAUCGAUCGAUGCUACGAUAGCUUAGCGGAA

Genome decoding

Genome → RNA sequence

The genotype and phenotype are embodied in different entities

The Potts Model



Random movements of the border accepted or rejected depending on the total energy of the cell:

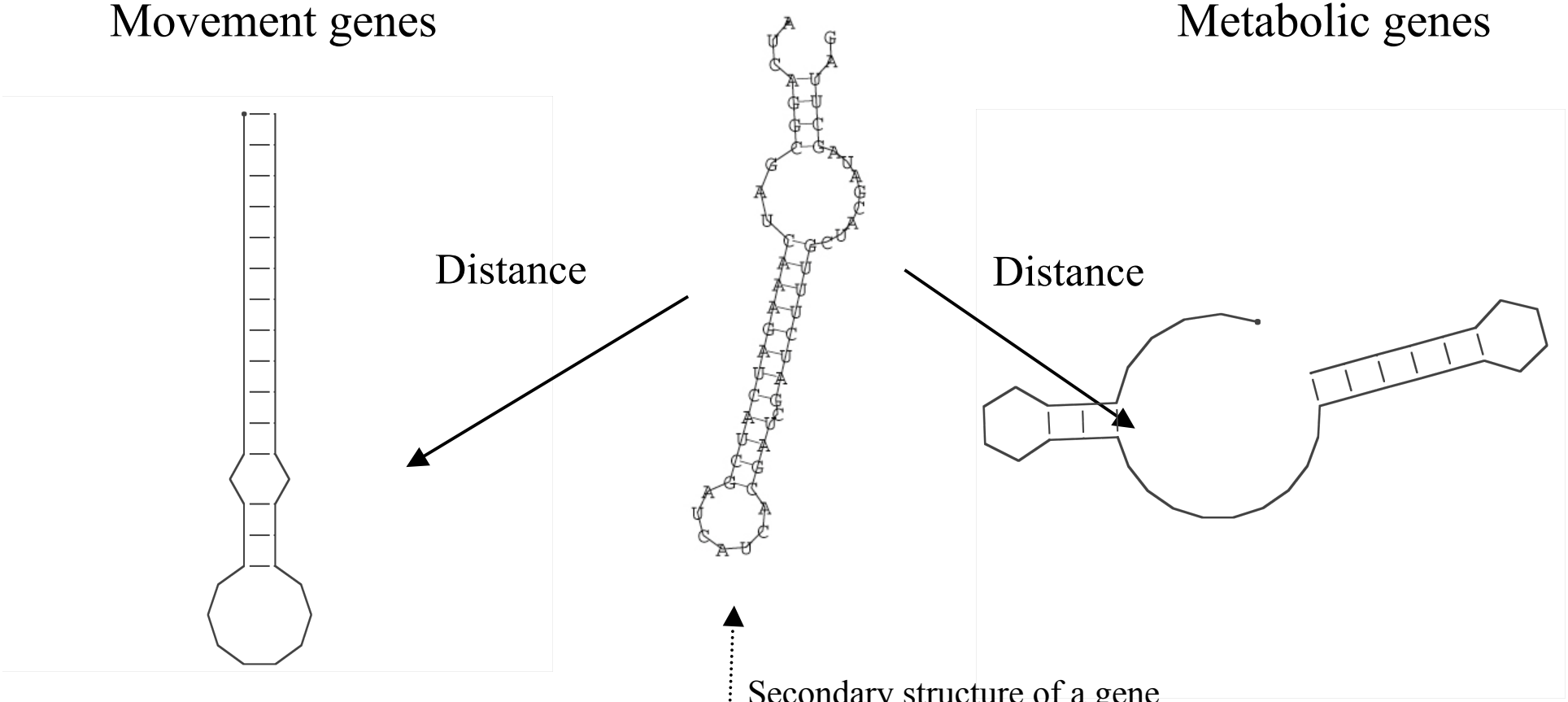
$$E_{cell} = \sum \frac{J_{type,type}}{2} + \sum J_{type,A} + \sum J_{type,S} + \lambda(v - V)^2$$

Over all entries of the membrane, where:
*-**J** is the interaction matrix between type of cells, cells and the air and cells and a substrate.*
*-**V** is the target volume and **v** the actual volume of a cell.*
*-**λ** is the compressibility of the cell.*

Genome and genes

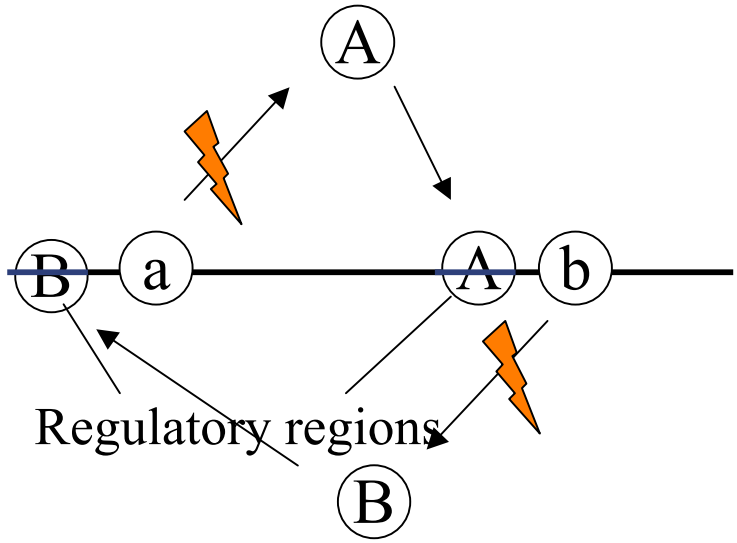
Movement genes

Metabolic genes



The regulatory network

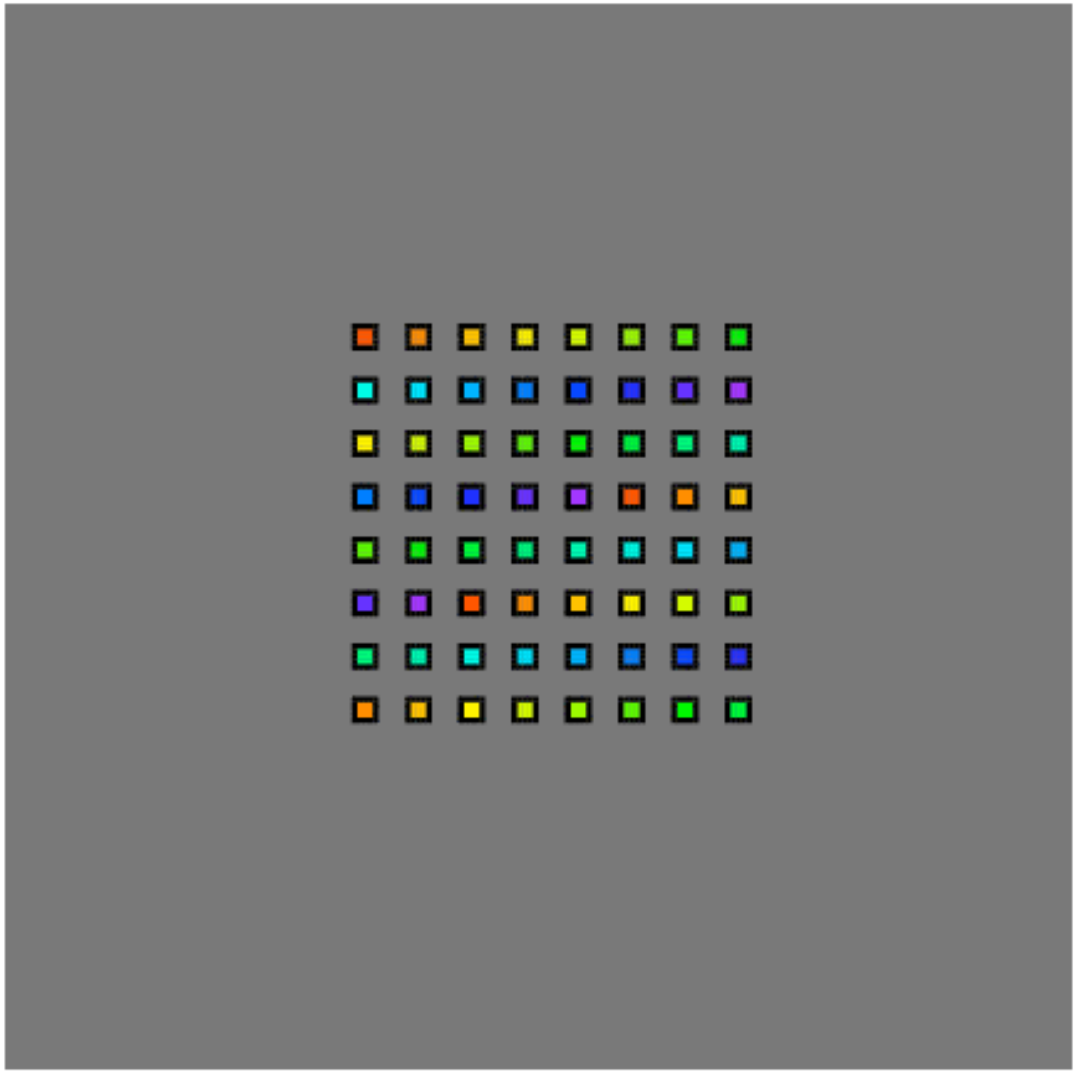
At the moment, we are using a very simple **regulation** network for our two kind of genes:



$$\frac{dp_m}{dt} = G_m \cdot k \frac{1}{1 + p_f^{\,3}} - c \cdot p_m$$

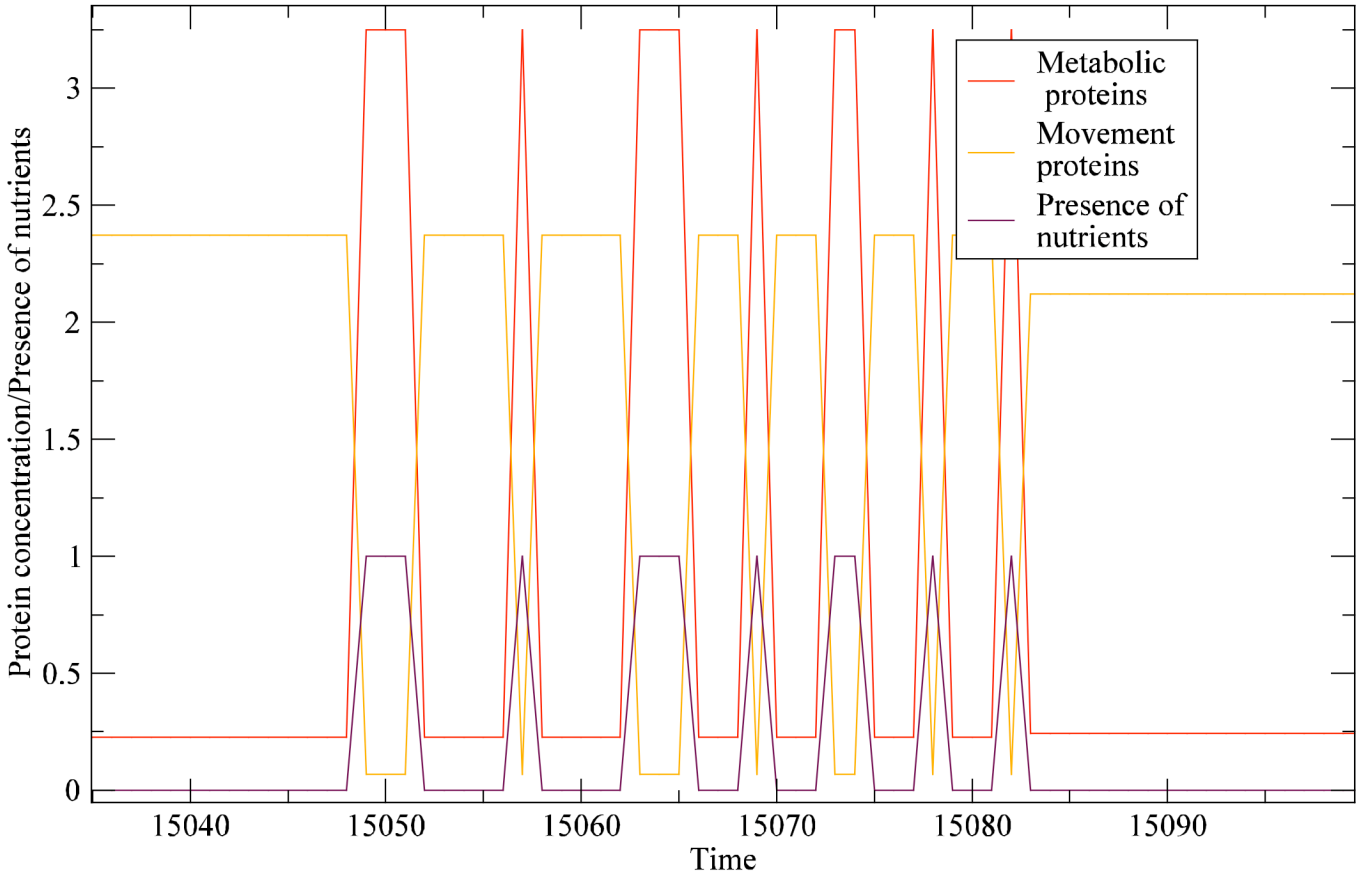
$$\frac{dp_f}{dt} = G_f \cdot k \frac{1}{1 + p_m^{\,3}} - c \cdot p_f$$

In motion

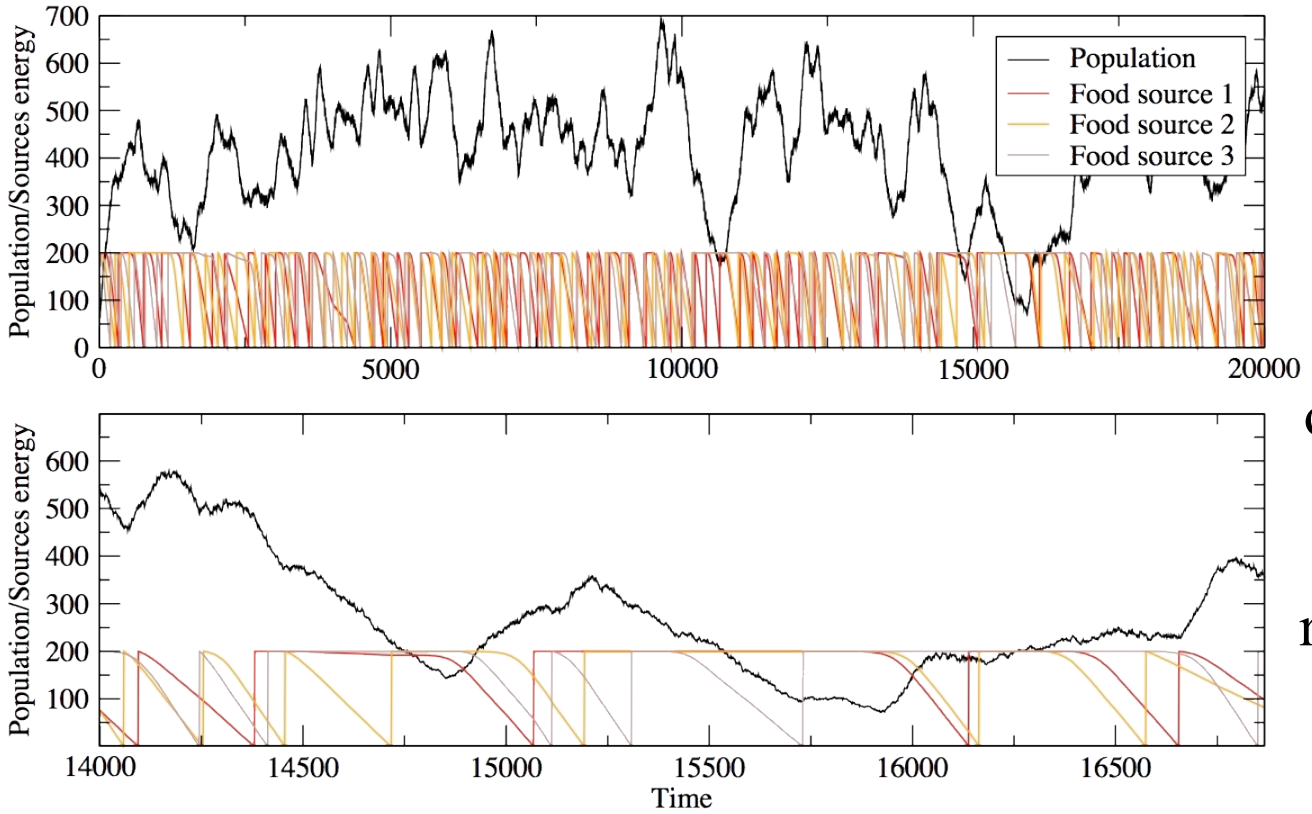


Protein expression curves

With these equations we obtain a switch between the **protein** concentrations every time an impulse is given from the outside.

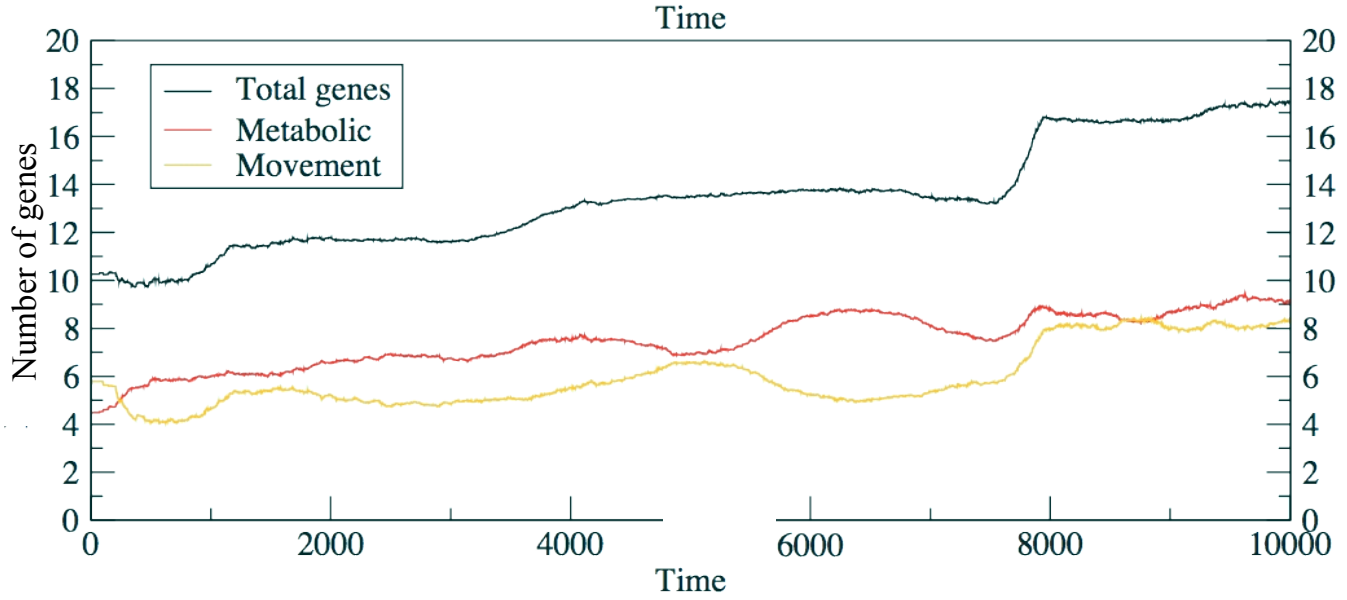


Population and food sources

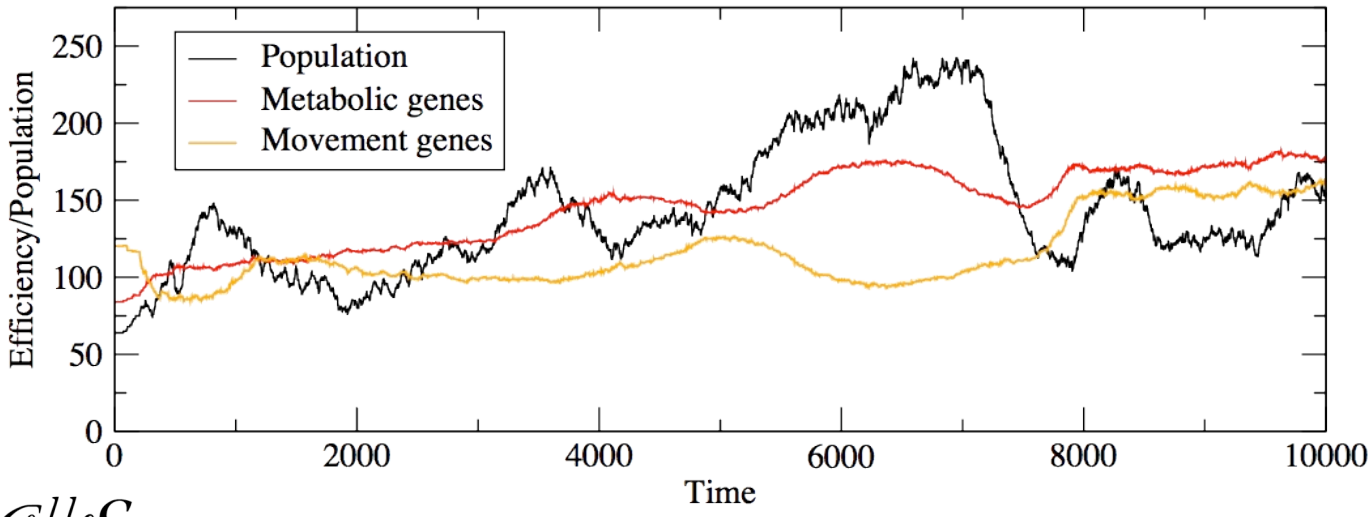


Population grows when cells are **feeding** from the **sources**. Depending on the changes in the environment, the number of cells is reduced or increased.

Genome evolution



Expected **number** of genes:
9.875
Number of genes at the end:
18



Difference between **efficiencies** is also controlled by the regulatory network

CelloS

Conclusions and outlook

Conclusions:

- Neutrality is decreased in the case of more than two sequences interacting.
- Depending on how interactions are defined, systems of several molecules can evolve as in models of individual sequences, the hypercycle for example, or be destroyed by the low neutrality of the map (results not shown).
- CelloS does not measure fitness as a real number but selection acts through surviving to changing conditions. (And selection does work)

Conclusions and outlook

Further work:

- More detailed study of cofold neutrality and its relation to the degree of connectivity of a network
- Decoding of the regulatory networks from the genome
- Phylogenetic trees from these models to be compared with real data

The end

I would like to thank:

Peter Stadler
Peter Schuster
Christoph Flamm
all the people of the TBI

and the audience