Molecular Computing: the activities of the MolCoNet Network

Giancarlo Mauri

Università di Milano-Bicocca

Network Objectives

"The aim is to stimulate the research on molecular computing, ..., through the coordination of the scientific activity of some of the major research groups active on this subject in Europe.

The WG will foster joint work and the exchange of results and knowledge among the involved research groups ....

Furthermore, it will assess the state-of-the-art and the key scientific and technological challenges in the area of bio-molecular computing, through the preparation and the regular updating of a roadmap report"

(Annex 1, Summary)

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Activities

- A yearly general meeting
- Short visits from one group to another
- Workshops on specific topics
- Organization of web pages on specific aspect of biological computing, including annotated bibliography, news, etc.
- Preparation and regular updating of a roadmap document, reporting on the state-of-the-art and the research directions
- Publication of papers in international journals and conferences

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Bio-molecular Computing

1994

Leonard Adleman solves HPP with biological techniques

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Hamiltonian Path Problem (HPP)

Instance: $< G, v_o, v_f >$ G=(V,E) digraph

Question: there a path from v_0 to v_f that enters every vertex exactly once?

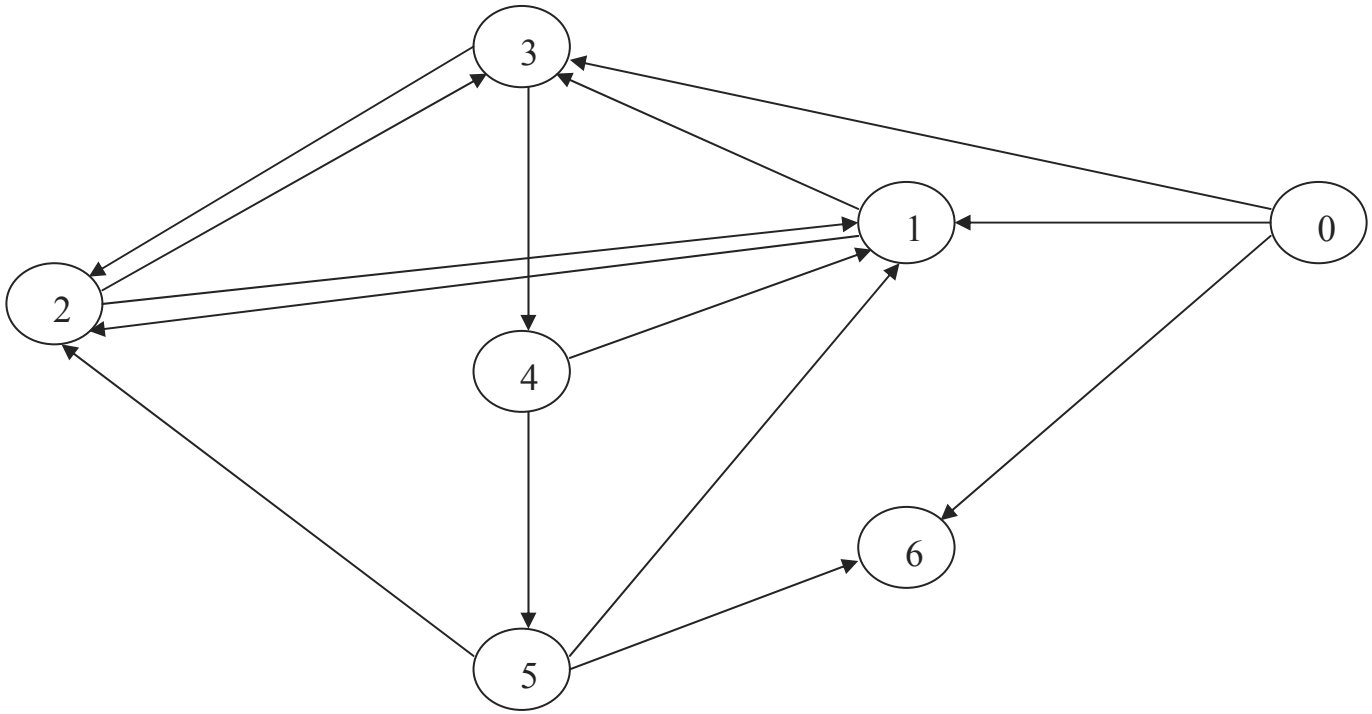
Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Adleman's experiment (1994)

- Solution of the hamiltonian path problem (NP-complete):
  - Graph with 7 nodes
  - Encoding with oligonucleotide sequences
  - Manipulation with standard molecular biology techniques
  - Proves the feasibility of molecular computations, even if for a specific problem and a "toy" instance

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The instance of the problem



There exists a path from 0 to 6 which goes exactly once through each node ?

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The encoding

$\Sigma = \{A, T, C, G\}$ set of nucleotides

$i \in V \longrightarrow O_i \in \Sigma^{20}$ (random choice)

Ex.

O_2 TATCGGATCGGTATATCCGA

O_3 GCTATTCGAGCTTAAAGCTA

O_4 GGCTAGGTACCAGCATGCCT

$(i,k) \in E \longrightarrow O_{i,k} = O_i(11,20)\ O_k(1,10) \quad (i \neq 0,\ k \neq 6)$

Ex.

$O_{2,3}$ GTATATCCGAGCTATTCGAG

$O_{3,4}$ CTTAAAGCTAGGCTAGGTAC

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The algorithm and the experiment

1. Path generation

- 50 pmol (about 3.10^{13} molecules) of cO_i for i = 1,…,5 and of $O_{i,j}$ for $(i,j) \in E$
- **ligase** reaction : cO_i is bound to $O_{i,j}$ (first half) and $O_{j,k}$ (second half) in double strand, guaranteeing the correct sequence of nodes

$O_{2,3}$ $O_{3,4}$

GTATATCCGAGCTATTCGAGCTTAAAGCTAGGCTAGGTAC

CGATAAGCTCGAATTTCGAT

cO_3

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The algorithm and the experiment

2. Selection of paths from 0 to 6

Polymerase chain reduction (PCR): cO_0 and cO_6 are used to bind molecules beginning with O_0 and ending with O_6 which are then duplicated

3. Selection of paths 7 nodes long

with agarose gel, the molecules 140 bases long are extracted

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The algorithm and the experiment

4. Selection of paths containing all the nodes

- Double strands are separated
- with cO_i ($i = 1, \ldots, 5$) molecules containing O_i are bound and selected

5. Answer

- with PCR and gel we verify if some molecule has been selected

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Bio-molecular Computing

- **Advantages of biological hardware**
  - **Massive parallelism**
  - **Low energy consumption**
  - **High information density**

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The other side of the moon...

- Errors in computation process

PCR Hybridization ...

To avoid this...

OPEN PROBLEM: Define suitable

ERROR CORRECTING CODES

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The other side of the moon...

- Lack of flexibility and generality

To avoid this...

Define suitable

FORMAL MODELS

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

The other side of the moon...

- Computational limits :
  - **"large" instances (500 nodes) remain intractable**
  - **"Weight" complexity**

To avoid this...

OPEN PROBLEM: Define suitable

ALGORITHMS

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Formal Models of DNA Computing

- Splicing Systems: they model the action of restriction enzymes on double stranded DNA molecules
- Insertion-Deletion Systems: based on biochemical operations that can be executed on single stranded DNA sequences
- Membrane Systems: in a hierarchical system of (cell) membranes, objects (chemicals) are modified following given rules

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Membrane Computing (P-systems)

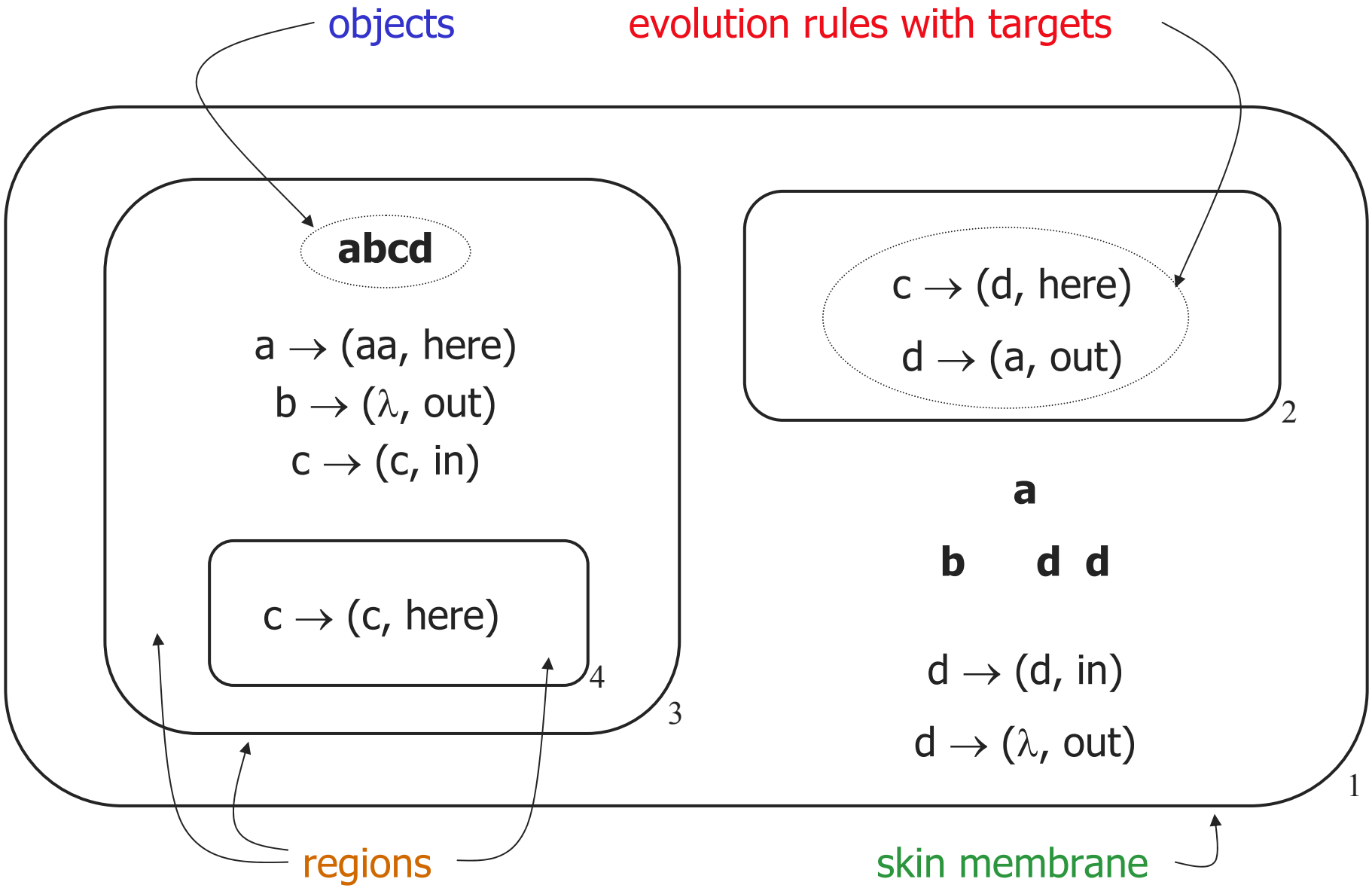
- P systems are a class of distributed and parallel computing devices (G. Paun, 1998) inspired by the structure and the functioning of living cells
- The main idea: making use of the way nature "*computes*" at the cellular level, where many processing of substances, energy and information take place
  - **biochemical substances inside the cell ↔ objects (symbols or strings over a given alphabet)**
- The basic model consists of a membrane structure: several cell-membranes are embedded inside an exterior membrane (the *skin membrane*)
  - **cellular processes and passage of substances through membranes ↔ evolution rules with target indications**

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Membrane Computing (P-systems)

- Each membrane delimits a region and it constitutes the communication channel between adjacent regions
- The membranes can contain objects, which evolve according to given evolution rules
- All rules are applied in a nondeterministic and maximally parallel manner: *all* the *objects* which can evolve should evolve, at each step and in each membrane of the system
- Starting from an initial configuration and letting the system evolve, we obtain a computing device

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

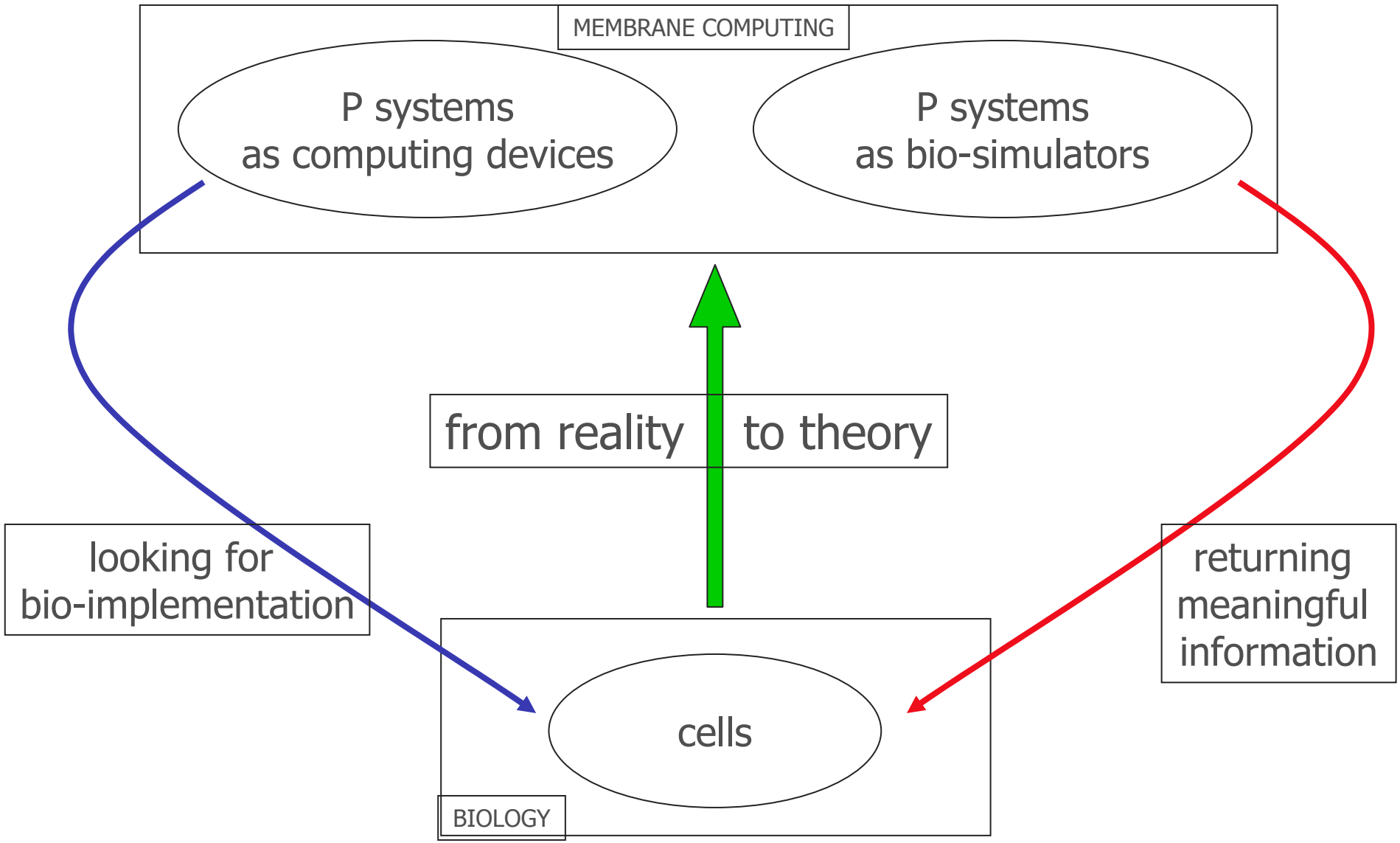

objects
evolution rules with targets
abcd
a → (aa, here)
b → (λ, out)
c → (c, in)
c → (c, here)
4
3
c → (d, here)
d → (a, out)
2
a
b d d
d → (d, in)
d → (λ, out)
1
regions
skin membrane

Computing with membrane systems

- Initial configuration (membrane structure, multisets of objects and sets of evolution rules)
- Nondeterministic and maximal parallel application of evolution rules:
  - all membranes are simultaneously processed
  - all objects inside a membrane simultaneously evolve
- Successful computation:
  - no rule can be further applied
  - output $\rightarrow$ (multi)sets of objects collected inside a prescribed membrane or outside the system

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Membrane Computing and Biology



Towards systems biology

- Models of mechanosensitive channels in bacteria
- Models of p53 signalling pathway

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

DNA word design

- Problems / errors:
  - Mismatches (non specific hybridization) ➲ false positives
  - **Self-annealing ➲ inefficiency**
  - **No hybridization ➲ false negatives**
- **Solution**
  - **Design of DNA codes with combinatorial constraints (Baum 96)**

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

DNA word design: goals

- to guarantee success in hybridization between a codeword and its complement
  - minimum length k for stability
- to avoid “false positives”
  - no long subsequence common to different codewords (or any concatenation)
- to avoid self-hybridization (secondary structure)
- to increase termodynamical efficiency (similar melting temperatures)
- to satisfy the constraints with minimal length words

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

DNA word design: tools

- **Error correcting codes**
- **Definition of good notions of "distance"**
- Not only DNA computing
  - design of probes for DNA microarrays
  - tag/antitag systems, chemical libraries

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Gene assembly in ciliates

Ciliates: very ancient group of single cell organisms (two billions years)
genetically very rich – at least 10.000 different organisms
very successful organisms, present anywhere where there is water.

One reason behind their success may be the way that they organize their genetic material, called "nuclear duality" – a feature unique to ciliates. They have two kinds of nuclei of very different functionality:

- micronucleus, used for sexual exchange of DNA in cell mating
- macronucleus, that contains genes needed for cell maintenance and reproduction

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Gene assembly in ciliates

- both micronucleus and macronucleus are present in multiple copies
- genes in macronucleus are expressed “all the time”
- genes in the micronucleus are kept inactive. The micronucleus is used for storing DNA until it is needed in the process of sexual reproduction
- during reproduction, the micronuclear genome is converted into the macronuclear genome – this process is referred to as gene assembly. Gene assembly is the most intricate DNA processing known in living organisms.

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Gene assembly in ciliates

- **Theory of pointer rewriting systems**
  - **Ehrenfeucht, Harju, Petre, Prescott, Rozenberg**
    Computation in living cells: gene assembly in ciliates
    Springer, 2003

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Web pages

- MolCoNet page
  - www.disco.unimib.it/molconet
- P Systems page
  - psystems.disco.unimib.it
- DNA computing page
  - http://www.dcs.ex.ac.uk/~pf201/dna.html
- EMCC page
  - http://openit.disco.unimib.it/emcc/

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Roadmap report

- **Molecular Computing: achievements and perspectives**
- **Gene Assembly in Ciliates**
  - **discusses the computational nature of involved DNA processing performed by ciliates (single cell organisms)**
- **Computing by Splicing**
  - **discusses an early theoretical model based on the use of restriction enzymes**

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Roadmap report

- **Test tube distributed systems based on splicing and related operations**
  - **covers parallel variants of the Splicing model**

- **Word Design for Molecular Computing**
  - **discusses the design of molecules to be used in molecular computing experiments**

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Roadmap report

- **DNA Computing in vitro**
  - gives an overview of techniques used throughout the whole field of molecular computing
- **Integrated DNA Computing**
  - discusses technological aspects of structuring systems for DNA computing

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

Roadmap report

- **Intra- and Inter-cellular processes**
  - discusses engineering and construction of reliable in-vivo logic circuitry
- **Membrane Computing**
  - presents a theoretical model inspired by the functioning of membranes in living cells

Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme