Métodos de Inteligencia Artificial en Biología Computacional
Artificial Intelligence Methods in Computational Biology

José Santos
Universidad de A Coruña
jose.santos@udc.es

Computational Biology. Definition

Bioinformatics: Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

Computational Biology: The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.

Bioinformatics applies principles of information sciences and technologies to make the vast, diverse, and complex life sciences data more understandable and useful. Computational biology uses mathematical and computational approaches to address theoretical and experimental questions in biology.

NIH Working Definitions of Bioinformatics and Computational biology, 2000:

https://www.bisti.nih.gov/docs/CompuBioDef.pdf

The NIH Biomedical Information Science and Technology Initiative Consortium agreed on the previous definitions of bioinformatics and computational biology recognizing that no definition could completely eliminate overlap with other activities or preclude variations in interpretation by different individuals and organizations.

Computational Biology. Definition

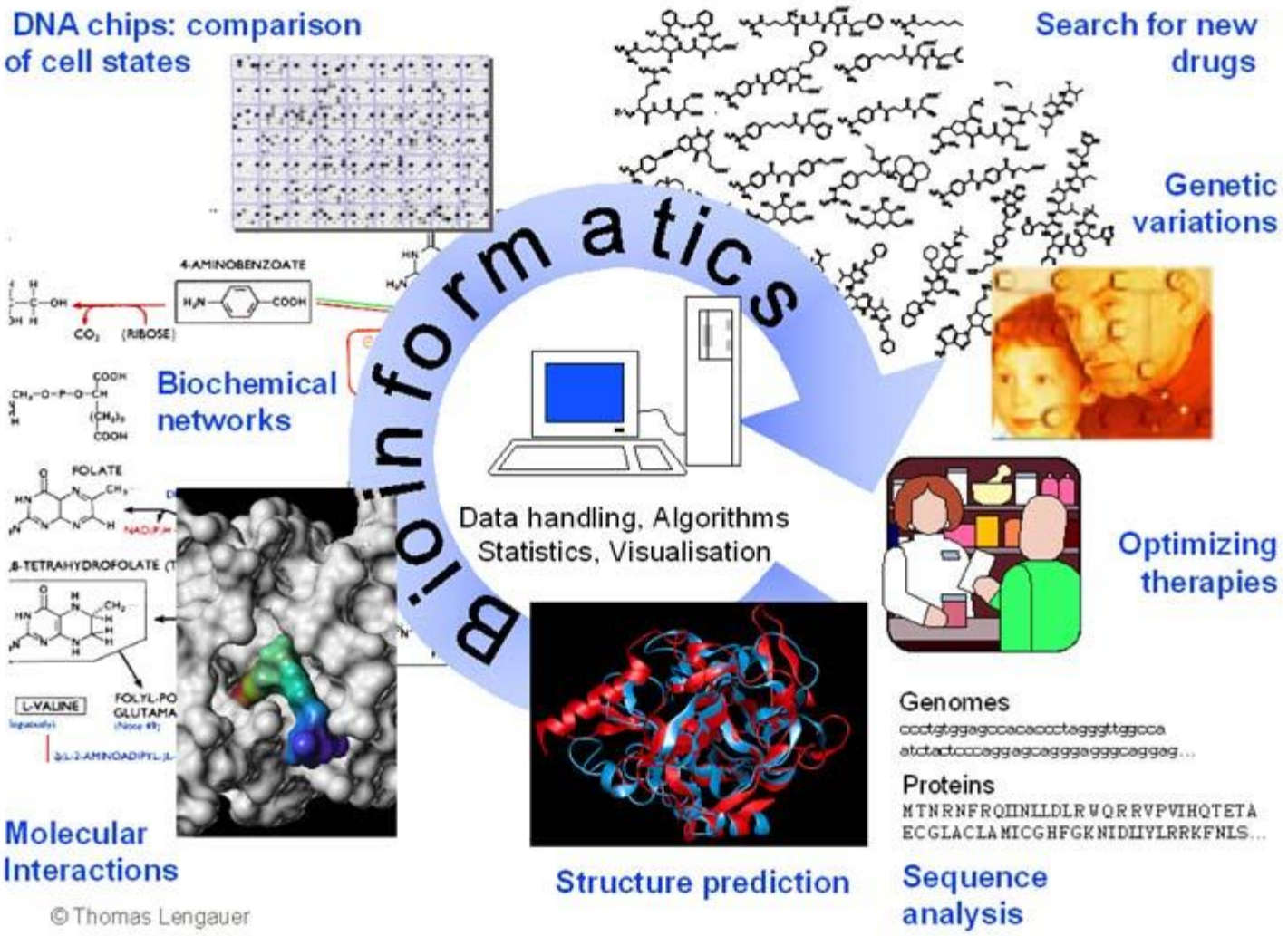
Other definition:

Computational biology: the study of biology using computational techniques. The goal is to learn new biology, knowledge about living systems. **It is about science.**

Bioinformatics: the creation of tools (algorithms, databases) that solve problems. The goal is to build useful tools that work on biological data. **It is about engineering**.

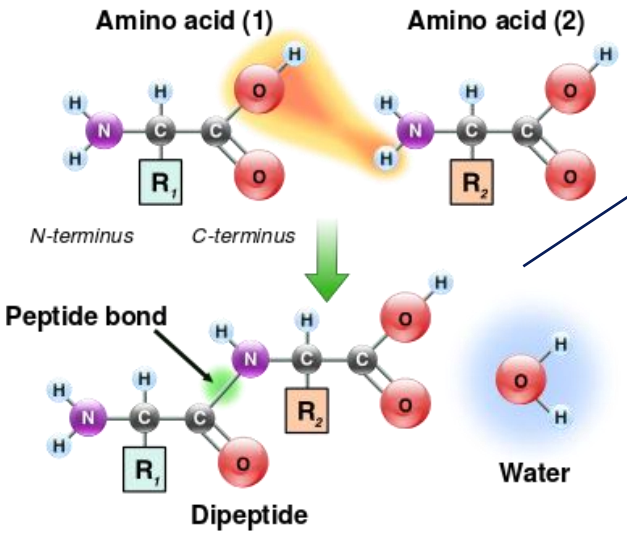
(Russ Altman, Stanford University).

Computational Biology. Areas of research

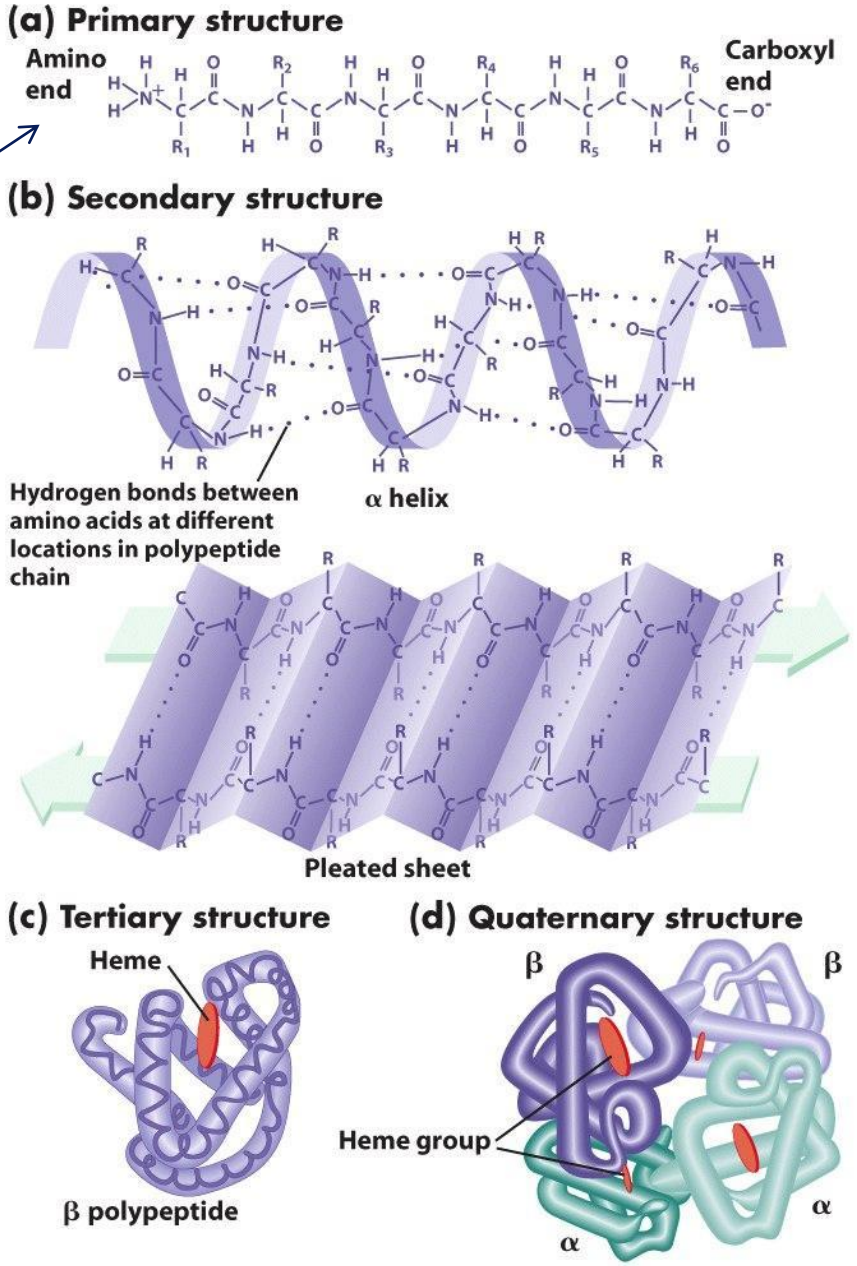


Protein structure prediction and protein folding modeling with cellular automata

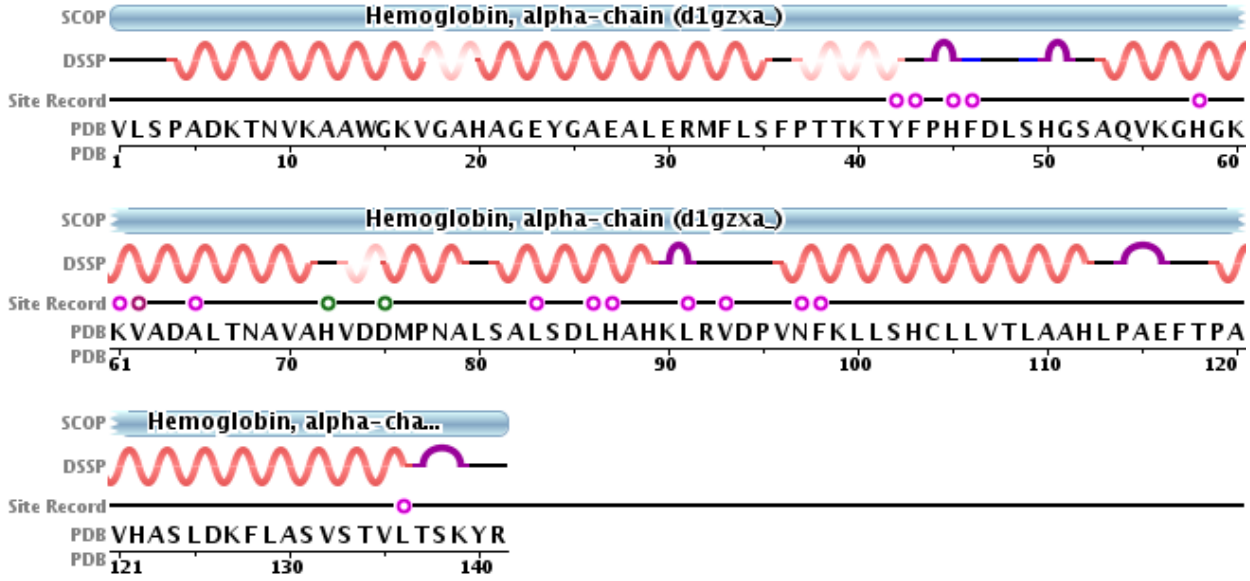
Proteins



"Ab initio" prediction



Proteins



Site Record Legend

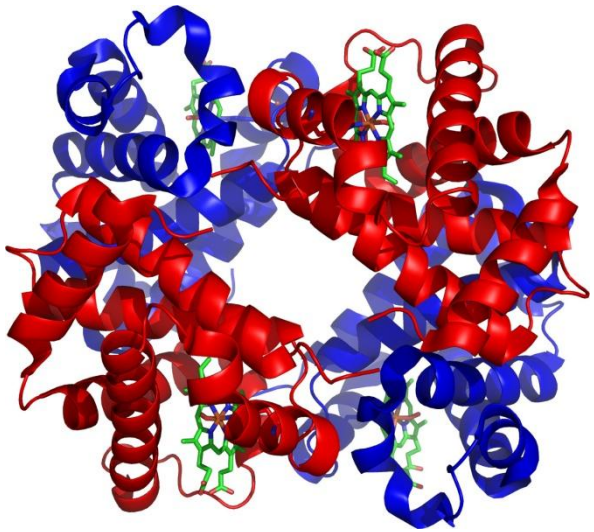
- BINDING SITE FOR RESIDUE OXY A1143 (SOFTWARE)
- BINDING SITE FOR RESIDUE HEM B1290 (SOFTWARE)
- BINDING SITE FOR RESIDUE HEM A1142 (SOFTWARE)

DSSP Legend

- T: turn
- empty: no secondary structure assigned
- G: 3/10-helix
- S: bend
- H: alpha helix

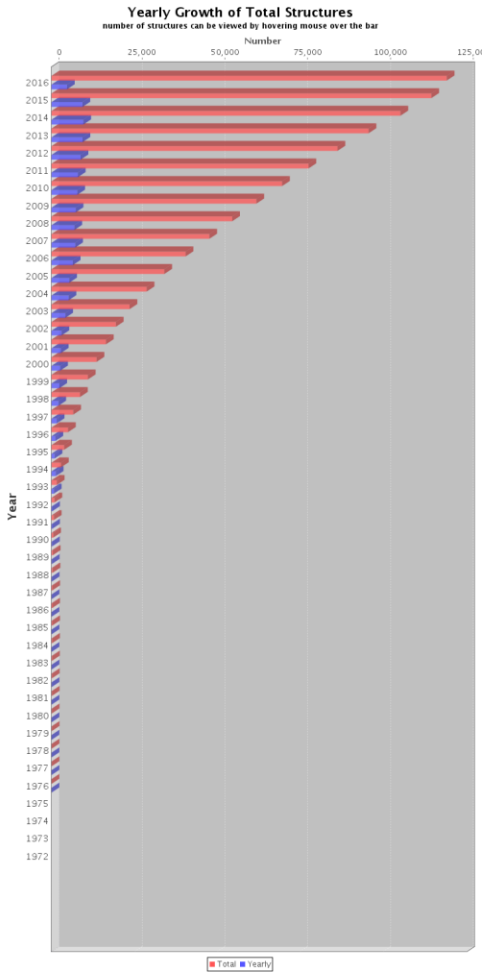
"Ab initio" prediction →

Protein Hemoglobin

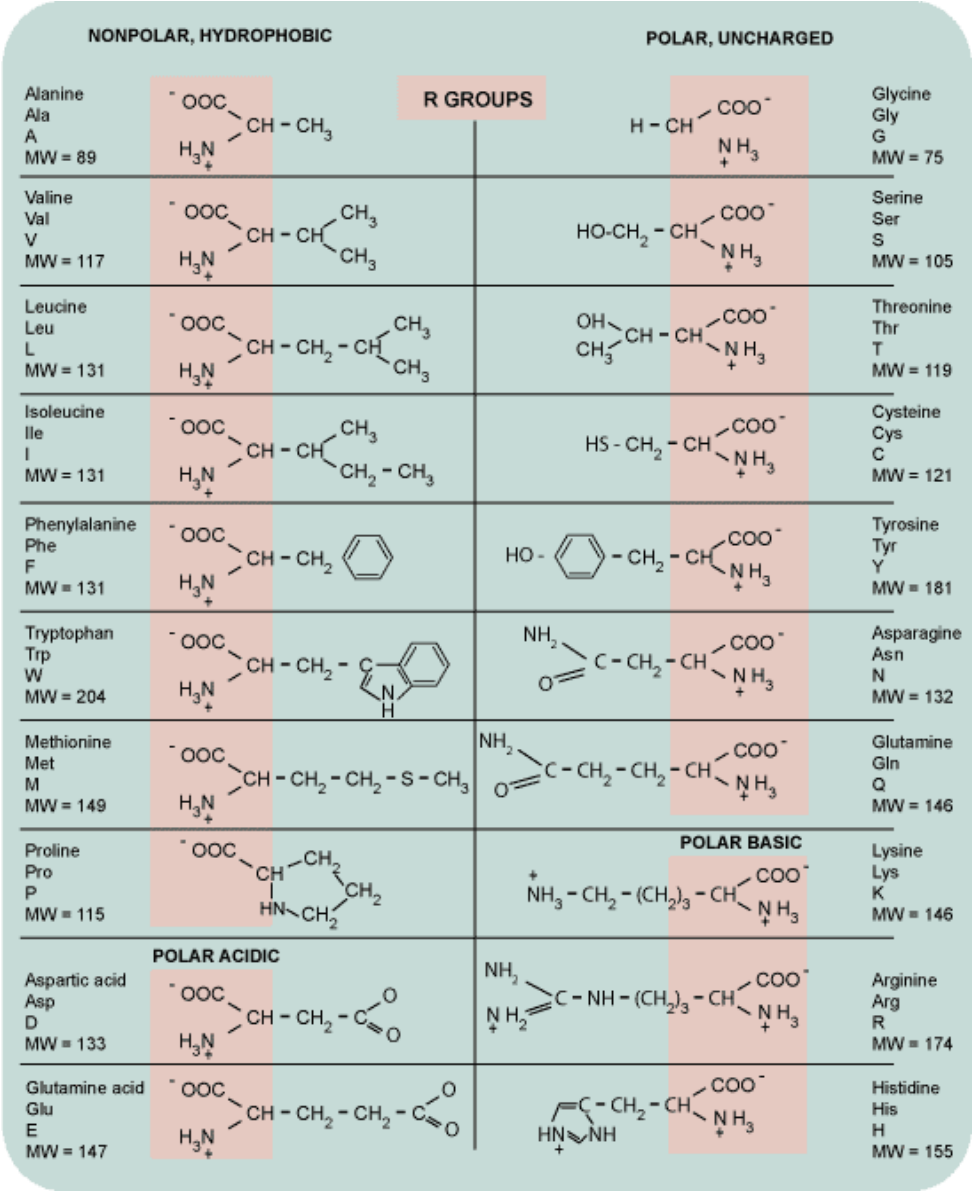


Proteins - PDB (Protein Data Bank)

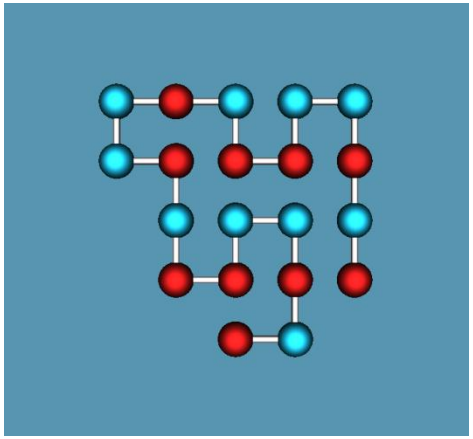
Experimental Method	Proteins	Nucleic Acids	Protein/Nucleic Acid complexes	Other	Total
X-ray diffraction	95636	1694	4817	4	102151
NMR	9840	1135	231	8	11214
Electron microscopy	666	29	227	0	922
Hybrid	83	3	2	1	89
Other	170	4	6	13	193
Total:	106293	2865	5283	26	**114569**



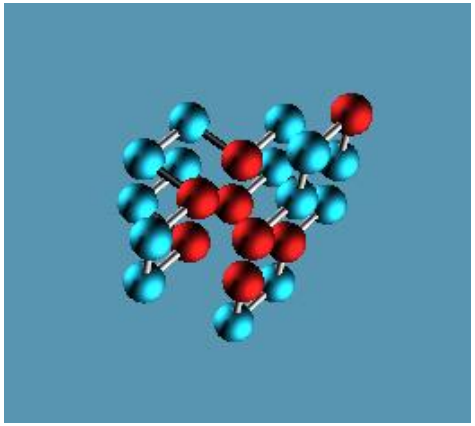
Proteins.
HP Model



HP model



Grid 2D

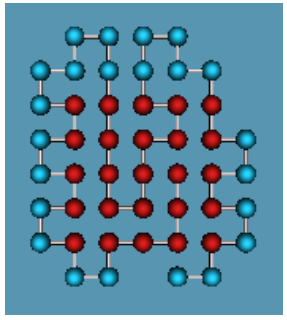


Grid 3D

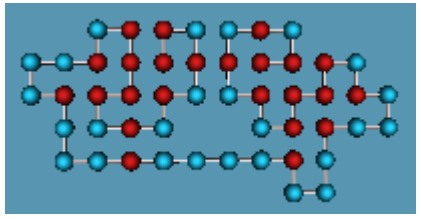
- Amino acids are classified in:
  - H (hydrophobic): low propensity to be in contact with water, tendency to be buried inside the protein core
  - P (polar): tendency to be in the protein surface in contact with water
- Each protein is represented as a chain:

  HPHPPHHPHPPHPHHPPHPH

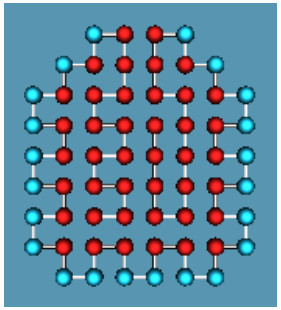
Benchmarks



Opt - 23



Opt- 21



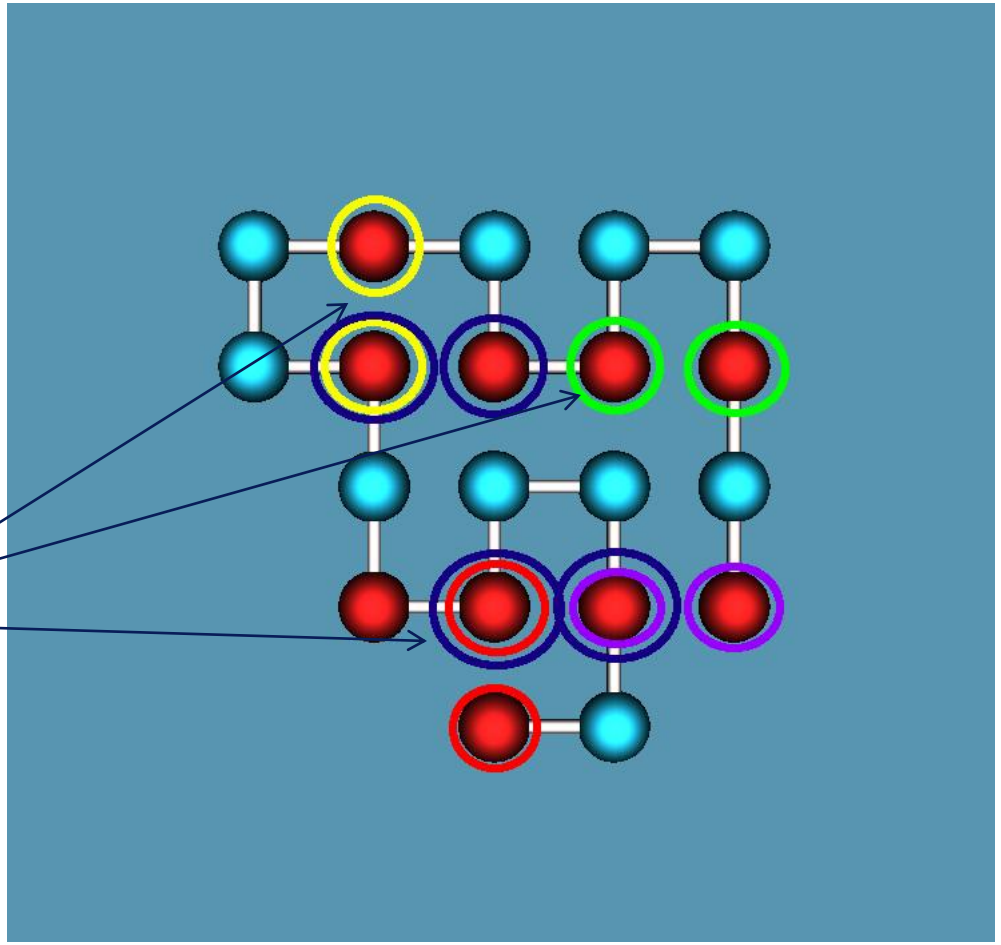
Opt - 42

HP Model. Protein Structure Prediction with Evolutionary Algorithms

HP energy matrix

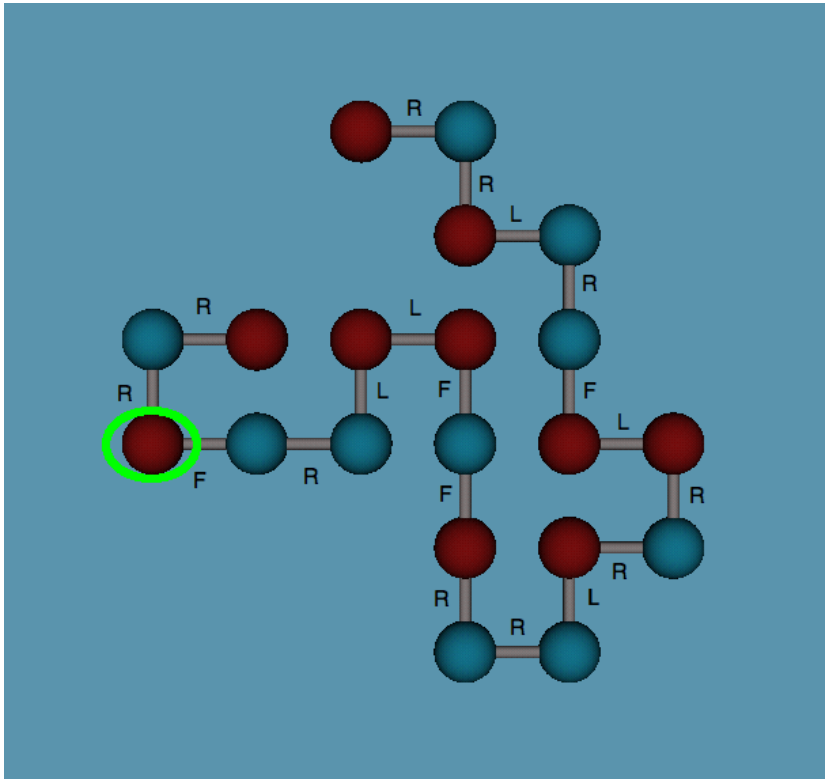
HH contacts

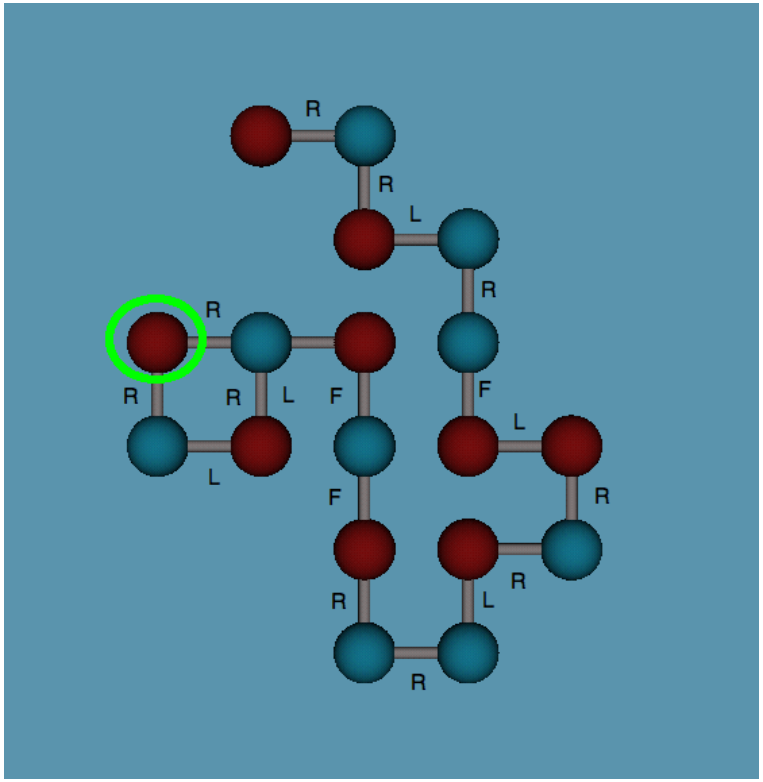
$$E = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$$



The energy reflects the fact that the hydrophobic amino acids have a propensity to form a hydrophobic inner core.

Protein Conformation Representation: relative moves





Relative moves: F, L, R (Grid 2D)

Differential Evolution

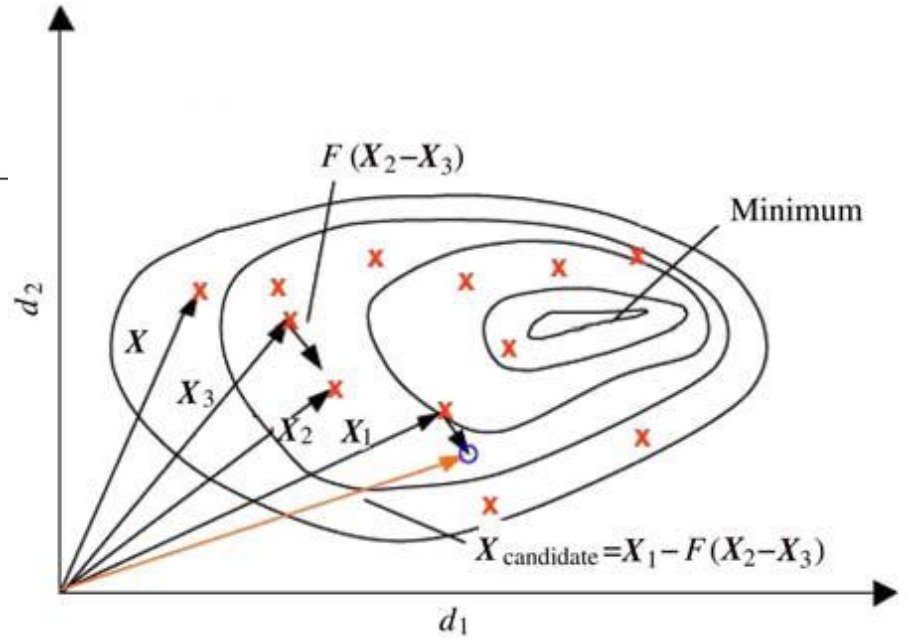
Algorithm 1 Differential Evolution Algorithm.

```
1: Initialize the population randomly
2: repeat
3:     for all individual x in the population do
4:         Let x1, x2, x3 ∈ population, randomly obtained {x1, x2, x3, x different from each other.}
5:         Let R ∈ {1, ..., n}, randomly obtained {n is the length of the chain.}
6:         for i = 1 to n do
7:             Pick ri ∈ U(0, 1) uniformly from the open range (0,1).
8:             if (i = R) ∨ (ri < CR) then
9:                 yi ← x1i + F(x2i − x3i)
10:            else
11:                yi = xi
12:            end if
13:        end for{y = [y1, y2...yn] is a new generated candidate individual}
14:        if f(y) < f(x) then
15:            Replace individual x by y
16:        end if
17:    end for
18: until termination criterion is met
19: return z ∈ population \∀t ∈ population , f(z) <= f(t)
```

In our application: F:0.9

CR:0.9

X1 selected with tournament



Differential Evolution. Encoding

We used relative coordinates. Three movements in 2D: (F)orward, (R)ight and (L)eft

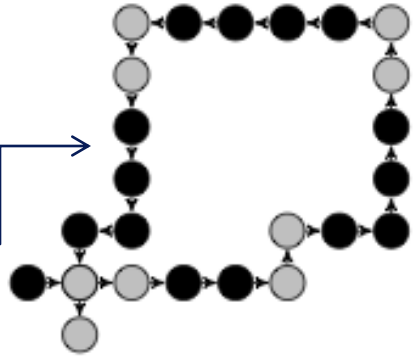
0.24	**-0.33**	**2.44**	**-1.25**	**0.18 …**
F	**F**	**R**	**L**	**F**

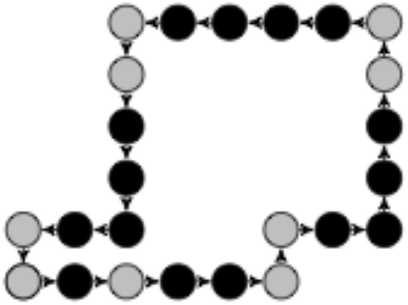
movement **L** if X_{ij} $[\alpha, \beta)$
F if Xij $[\beta, \delta)$
R if Xij $[\delta, \gamma]$

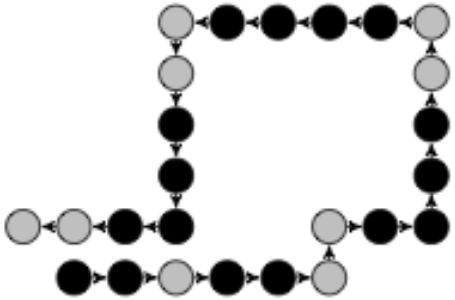
$\alpha < \beta < \delta < \gamma$ arbitrary constants in R (α=-3, β=-1, δ=1, γ=3)

Repair process

Absolute moves procedure: tries to maintain the relative conformation of the rest of the chain

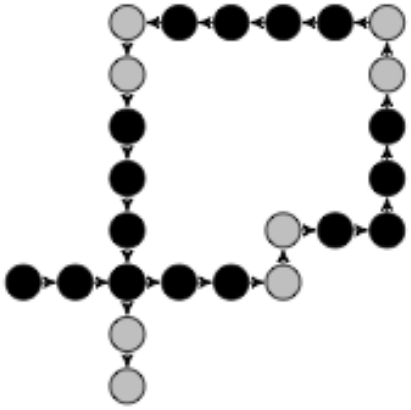






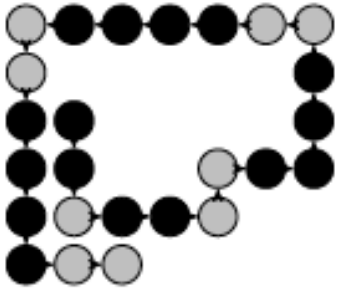
Intermediate steps

Final conformation



Original illegal conformation

Cartesian coordinates procedure: tries to obtain legal conformations searching for a similar one in the Cartesian space



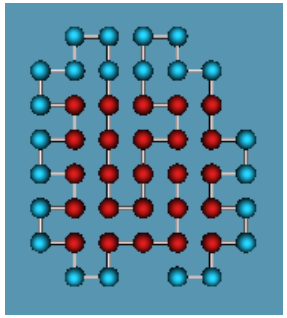
Final conformation

Some results

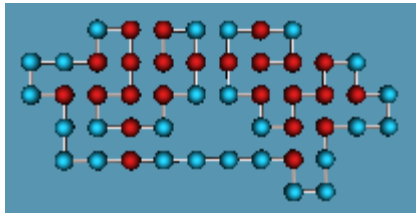
Comparison of results with the benchmark sequences

Seq.	E_{min}	U&M GA [23]	hybrid DE	L&B DE [13]	ACO [18]
S1	-9	**-9** (30492)	**-9**,-9 (3584, 6362)	**-9**,-9	**-9**
S2	-9	**-9** (30491)	**-9**,-9 (5806, 9292)	**-9**,-9	**-9**
S3	-8	**-8** (20400)	**-8**,-8 (7061, 18828)	**-8**,-8	**-8**
S4	-14	**-14** (301339)	**-14**,-14 (45793, 92579)	**-14**,-13.96	**-14**
S5	-23	-22 (126547)	**-23**,-23 (245943, 532787)	**-23**,-23	**-23**
S6	-21	**-21** (592887)	**-21**,-21(365222, 691989)	**-21**,-21	**-21**
S7	-36	-34 (208781)	-35,-33.57	-35,-34.79	**-36**
S8	-42	-37 (187393)	**-42**,-42 (176313, 340917)	**-42**,-41.87	**-42**

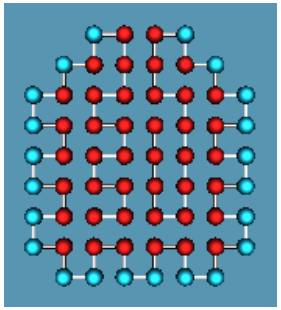
Optimal configurations for three of the sequences: S5, S6 and S8



S5 - 23

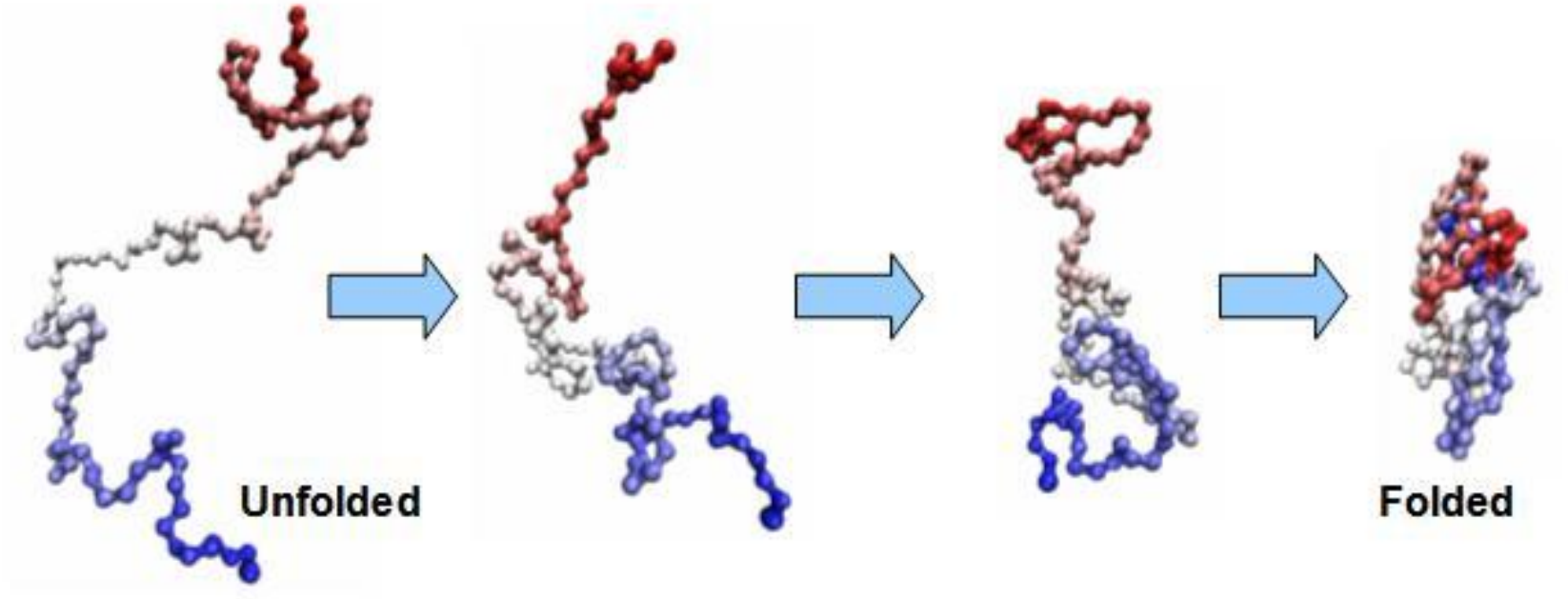


S6 - 21

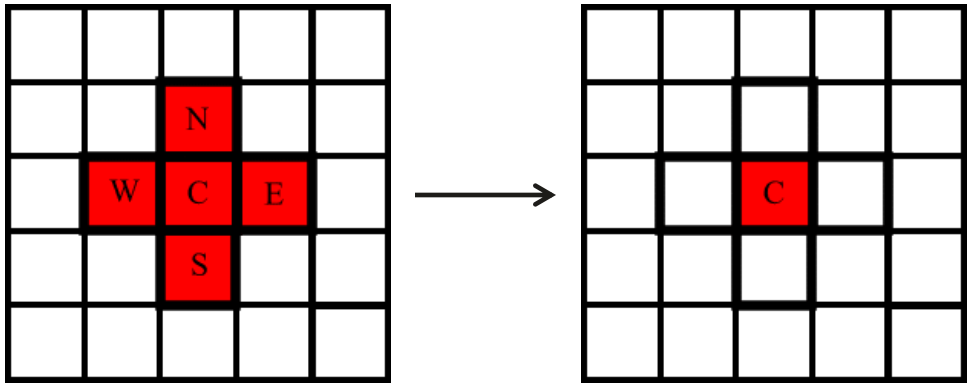


S8 - 42

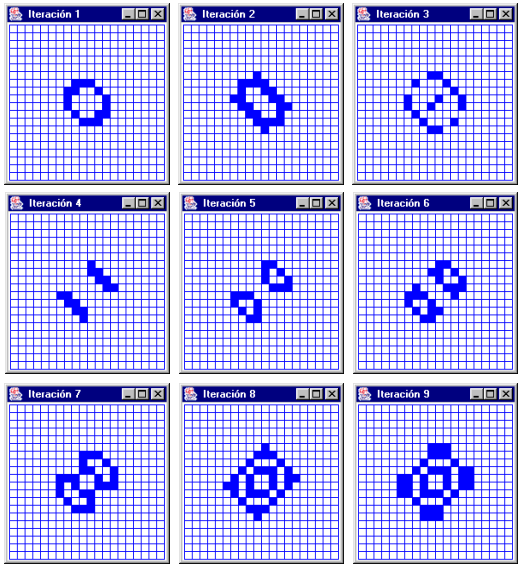
Protein folding



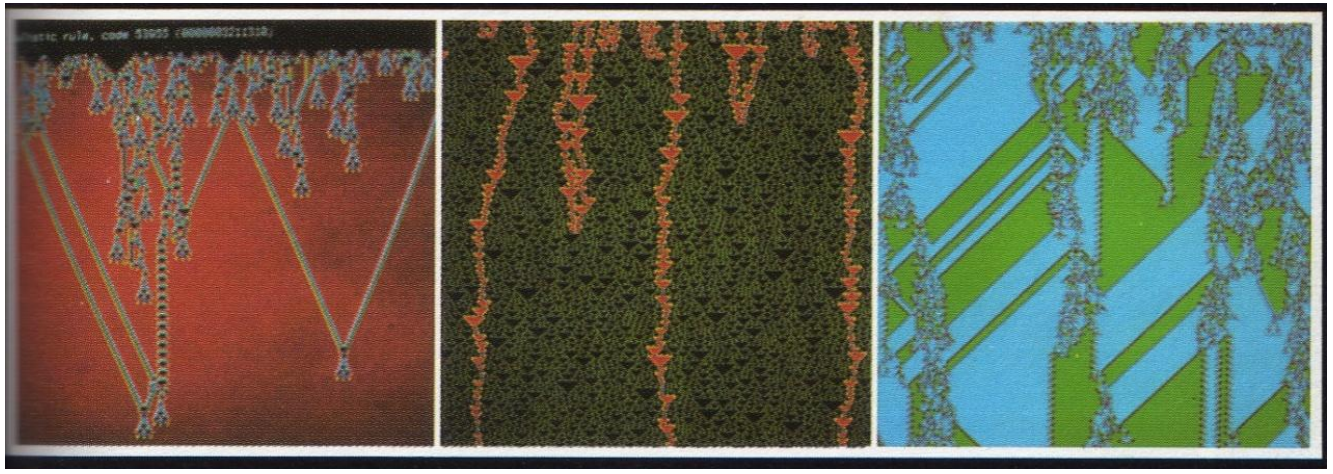
Cellular automata



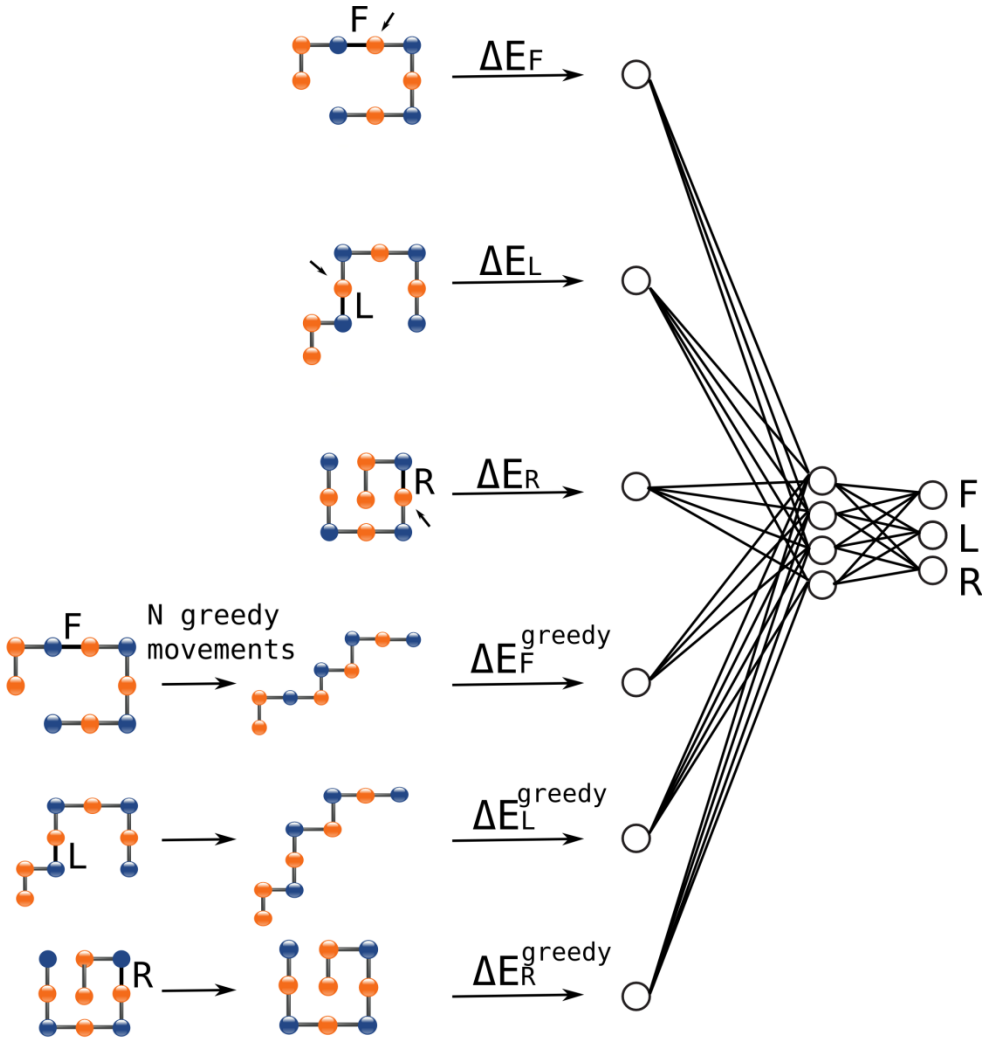
von Neumann neighborhood in a cellular automaton in 2D



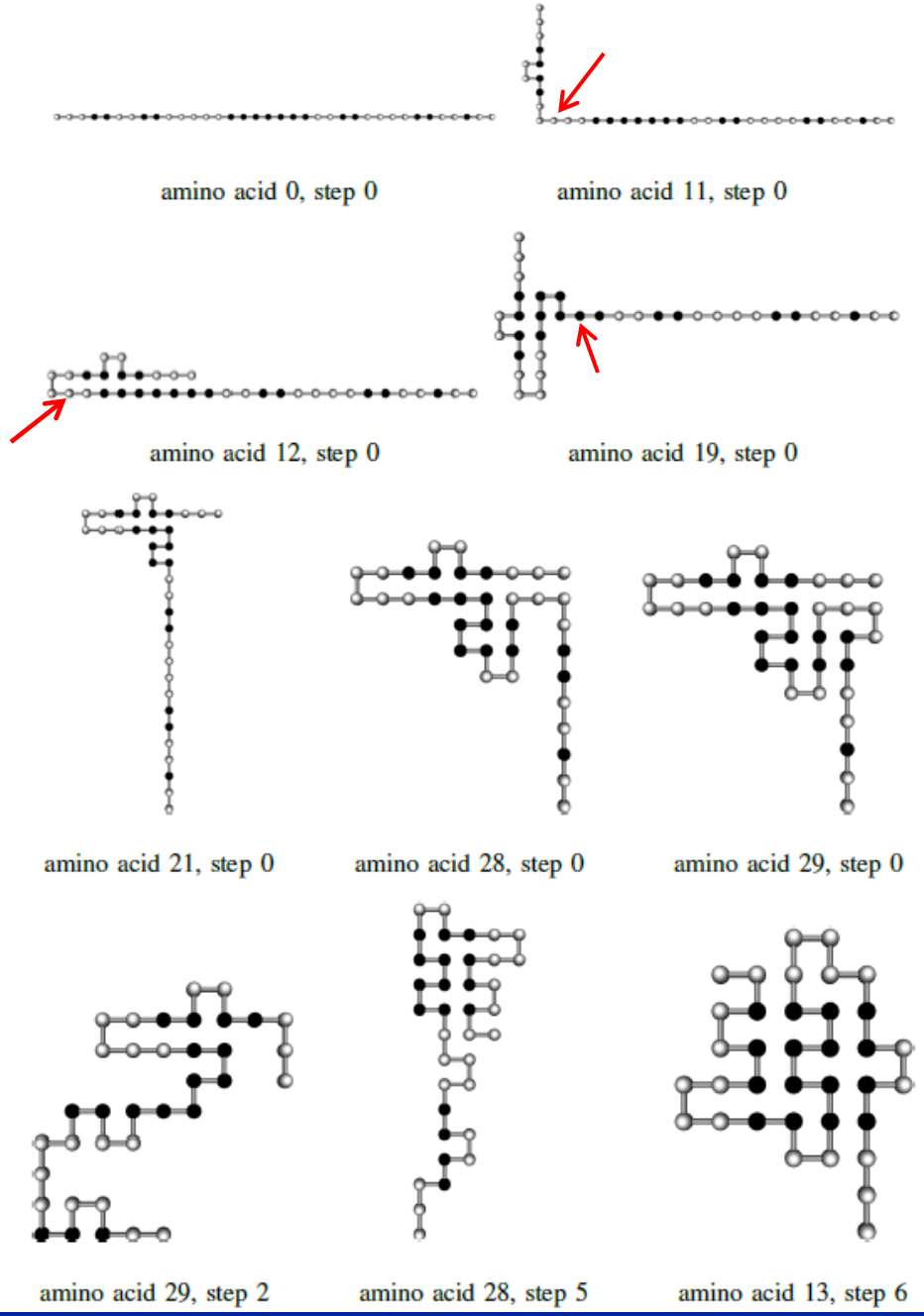
Conway's "Game of Life"



Protein folding with “Neural Cellular Automata”

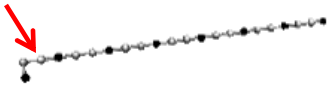


Protein folding. Example in 2D with the HP model



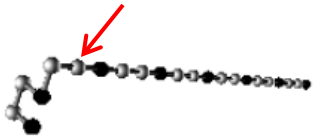
Protein folding. Example in 3D with the HP model

initial unfolded conformation

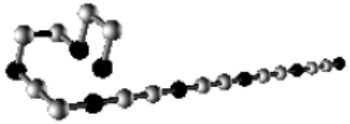


amino acid 2, step 1

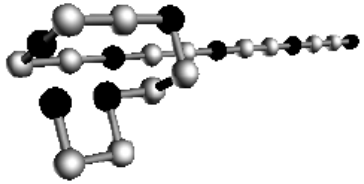
amino acid 3, step 1



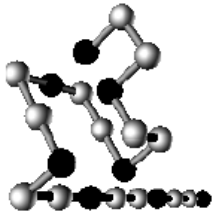
amino acid 5, step 1



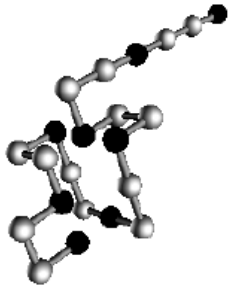
amino acid 8, step 1



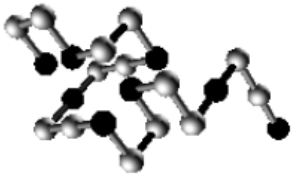
amino acid 9, step 1



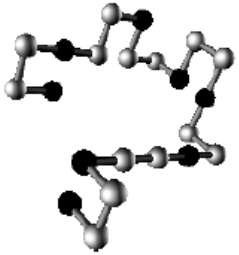
amino acid 11, step 1



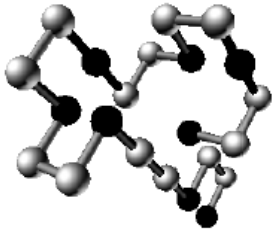
amino acid 13, step 1



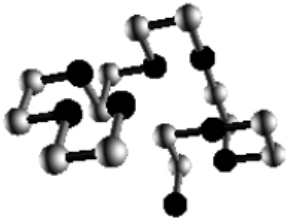
amino acid 15, step 1



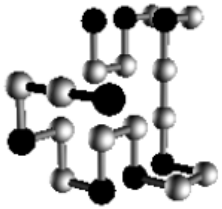
amino acid 13, step 2



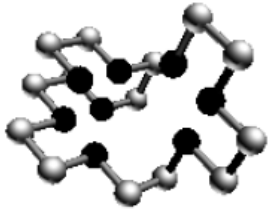
amino acid 21, step 2



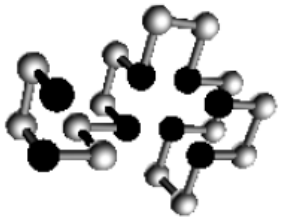
amino acid 20, step 3



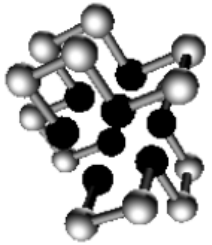
amino acid 5, step 4



amino acid 13, step 4

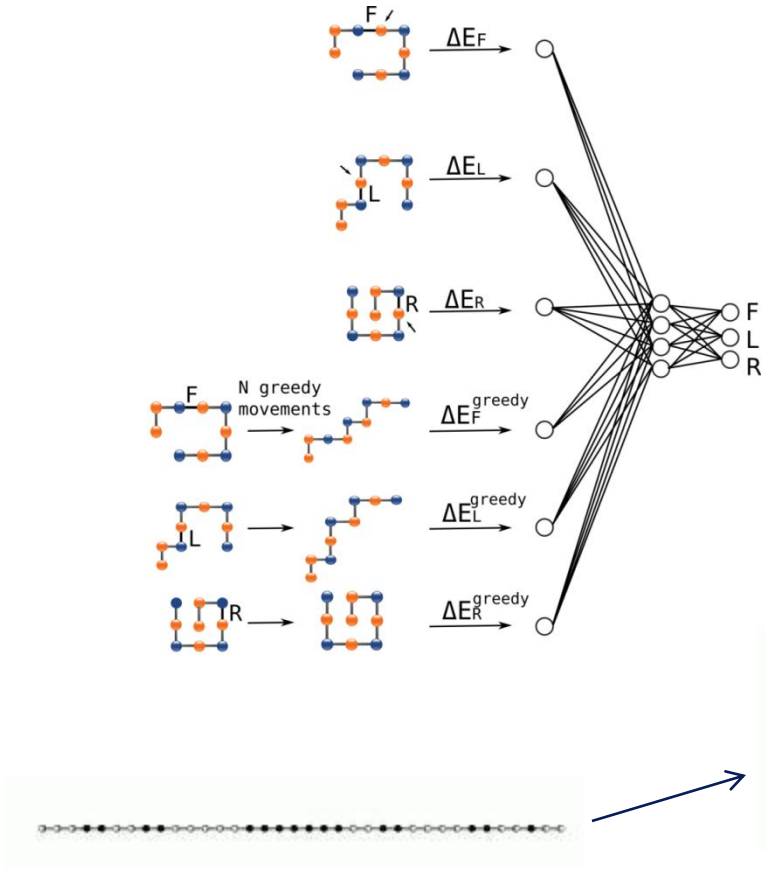


amino acid 4, step 5



amino acid 5, step 5

Evolving "Neural Cellular Automata" with Differential Evolution



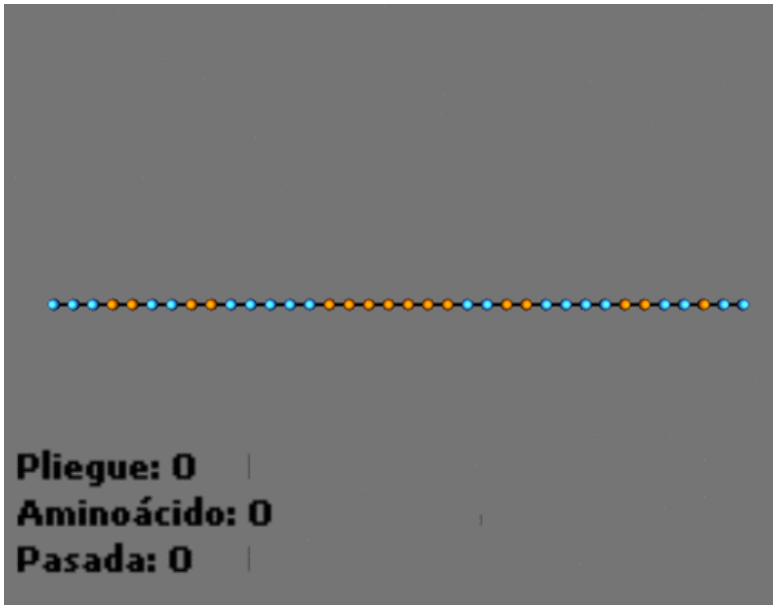
Genotype

Connection weights w_{ji}

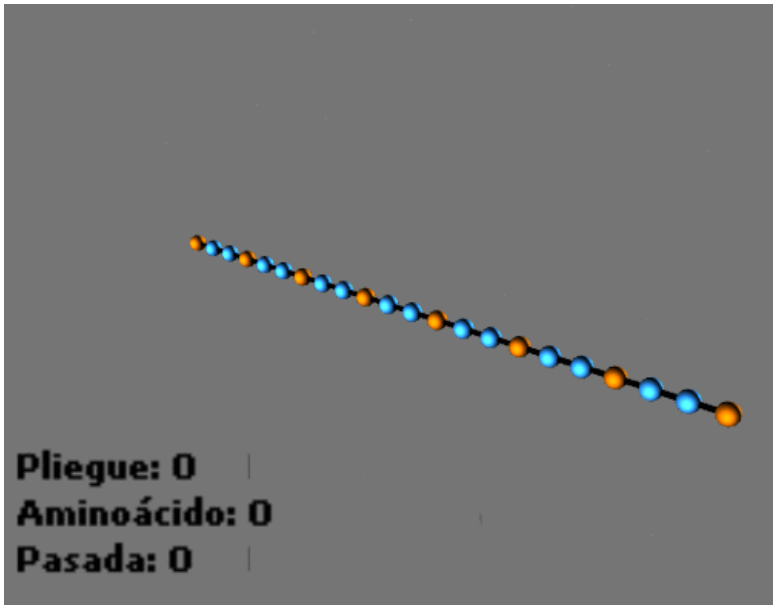
Fitness function

Number of final HH contacts

Protein folding process modeling with the HP model

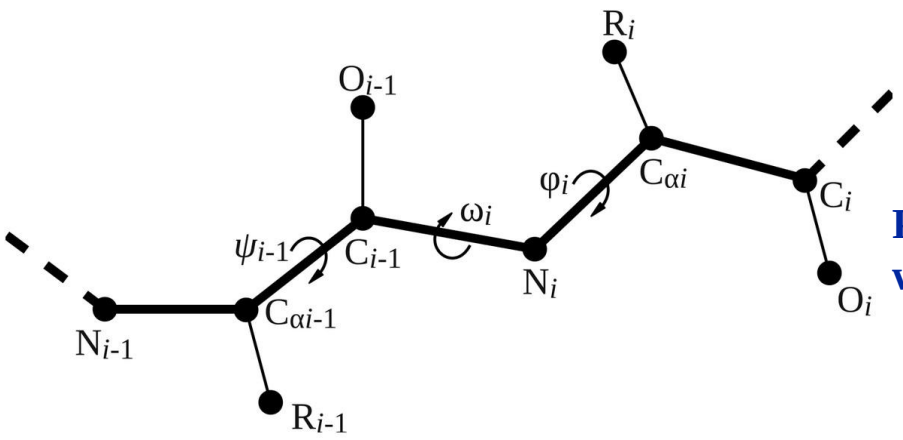


2D sequence – Optimum: 14 HH contacts

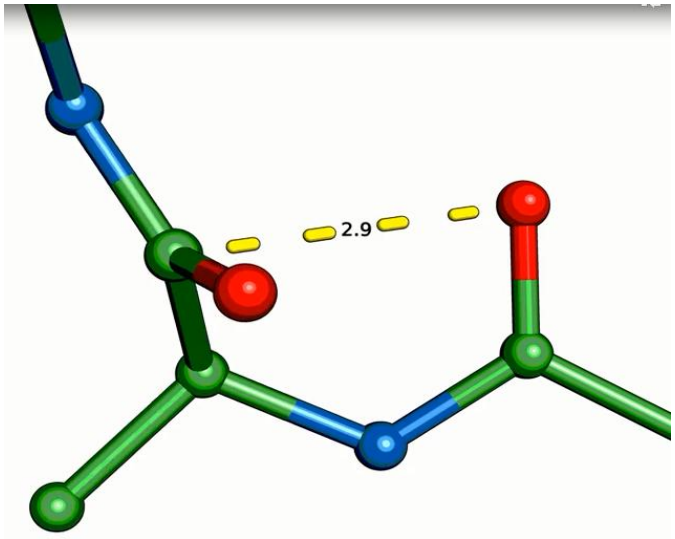


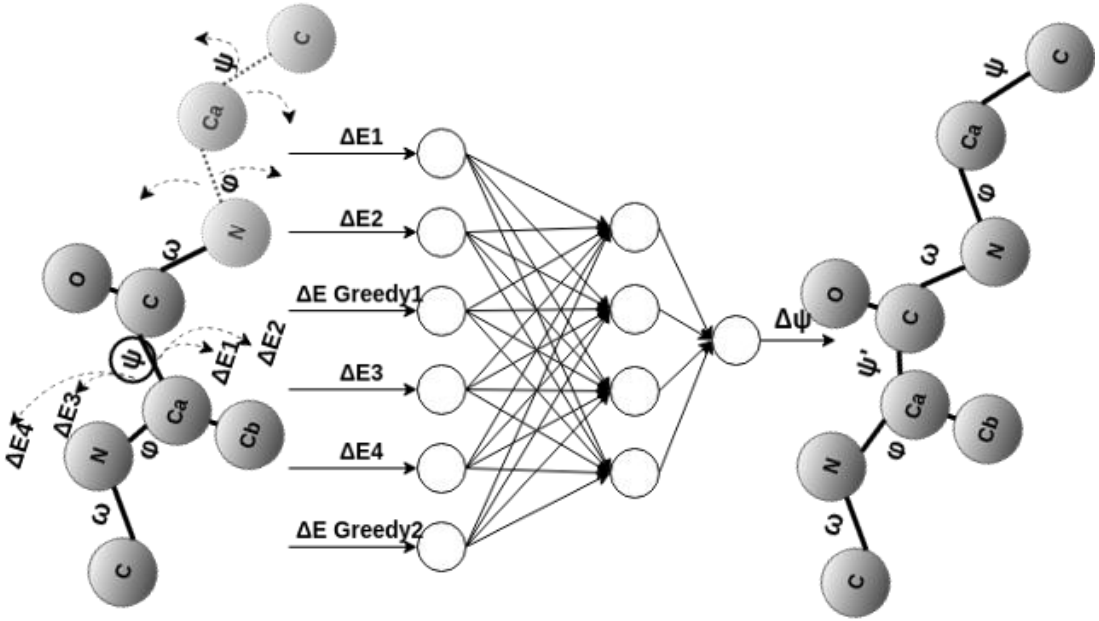
3D sequence – Optimum: 12 HH contacts

Protein folding process modeling with the off-lattice model of the Rosetta system

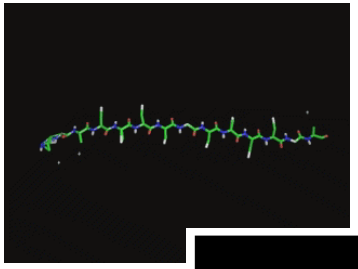


Protein conformation representation with the dihedral angles

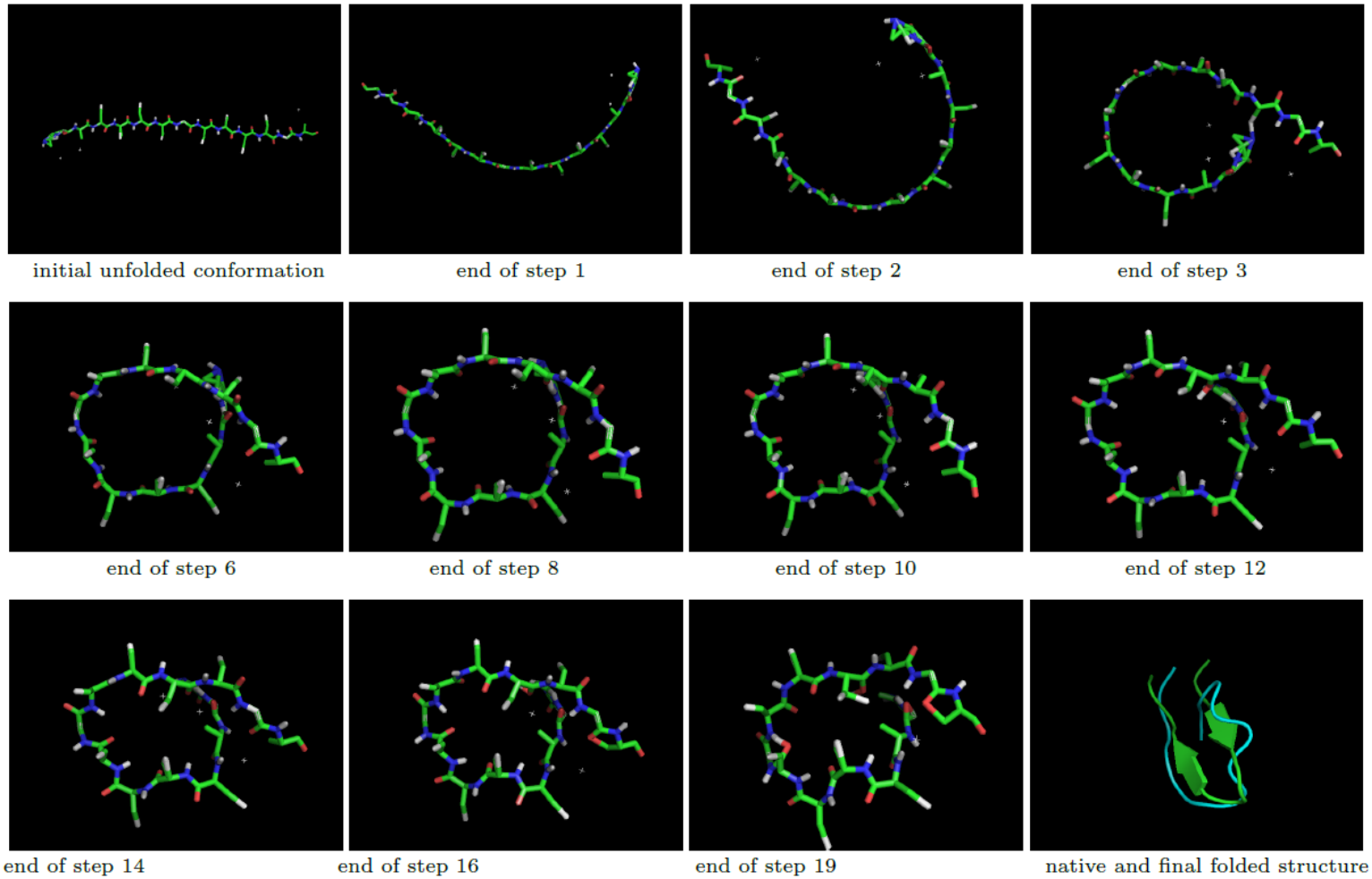




Neural cellular automaton

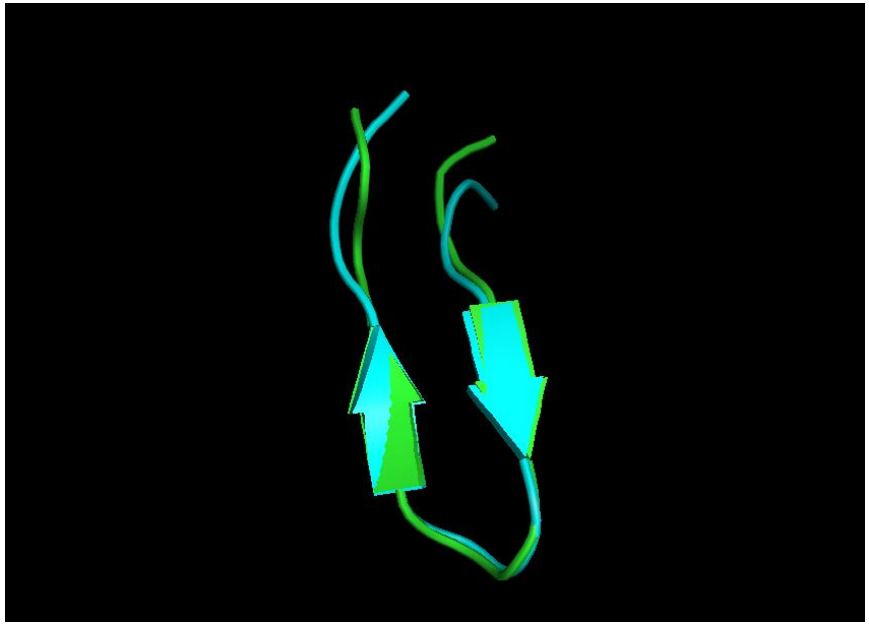


Protein folding process modeling with the off-lattice model of the Rosetta system

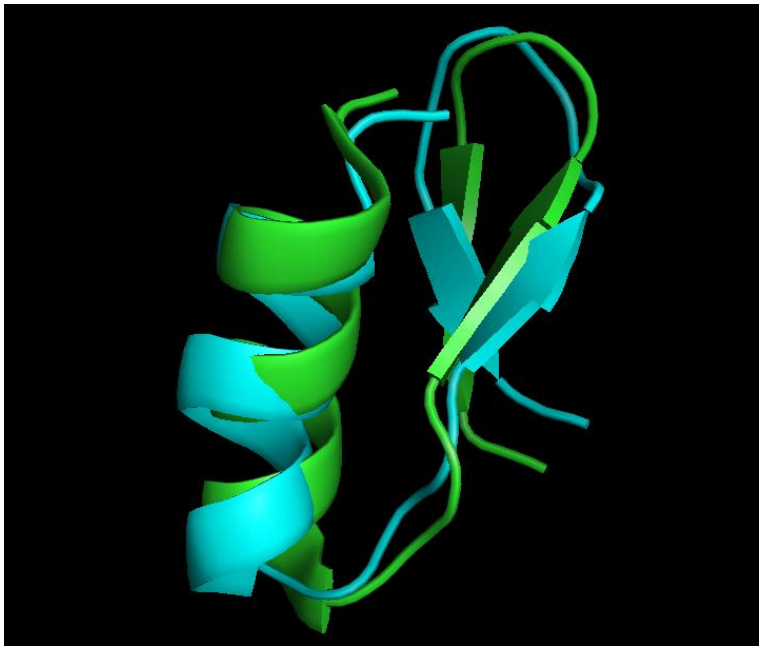


Snapshots of the folding at the end of different temporal steps with protein sequence 1j4m. The last snapshot shows the native structure (green) and the final folded structure at the last step (blue).

Protein folding process modeling with the off-lattice model of the Rosetta system



Final folded conformation for protein 1j4m (14 amino acids) (blue) and the corresponding native structure (green) when the energy term associated with secondary structure elements is taken into account and when the folding process begins with a partially folded conformation.



Final folded conformation for protein 1d5q (27 amino acids) (blue) and the corresponding native structure (green) when the folding process begins with a partially folded conformation.

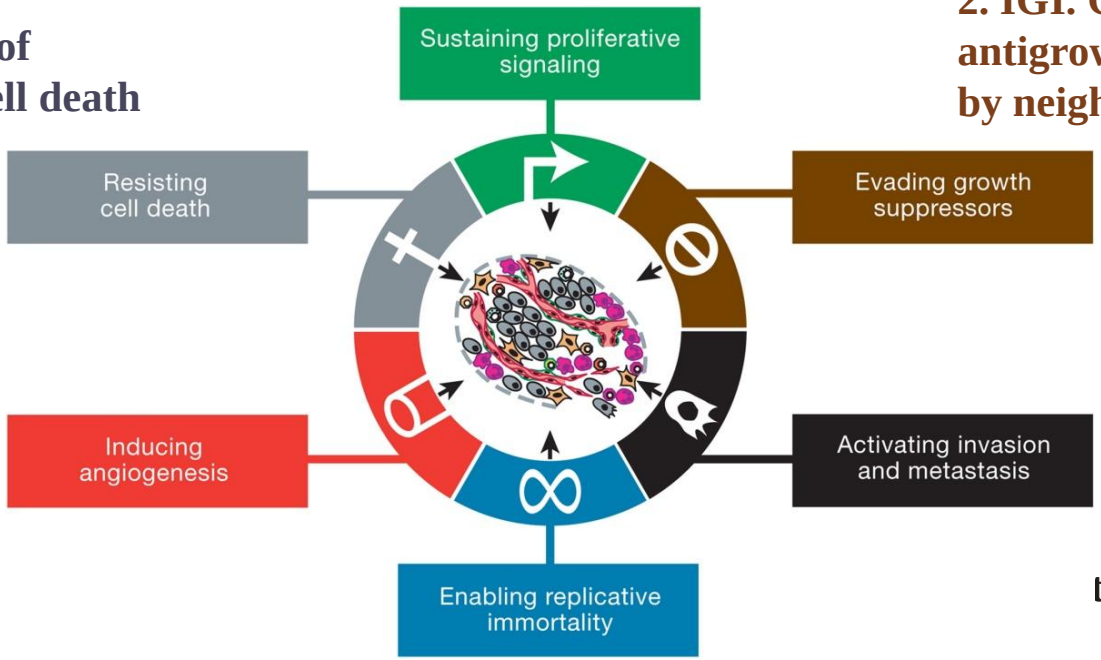
Tumor growth modeling with cellular automata

Cancer hallmarks

1. SG. Growth even in the absence of normal "go" signals.

2. IGI. Growth despite antigrowth signals issued by neighboring cells.

3. EA. Evasion of programmed cell death (apoptosis)



6. AG. Ability to stimulate blood vessel construction (angiogenesis)

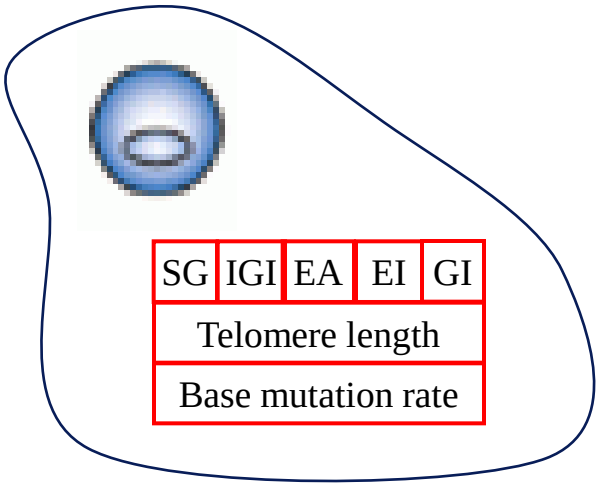
7. MT. Power to invade other tissues and spread to other organs.

4. EI. Effective immortality.

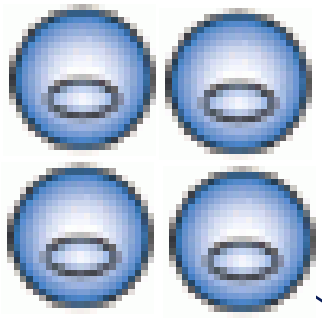
5. GI. Genetic Instability.

Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer, *Cell*, 100, 57-70.

AC Cell growth simulation



Cell “genome”



Mitosis event in the future: between 5 and 10 instant in the future

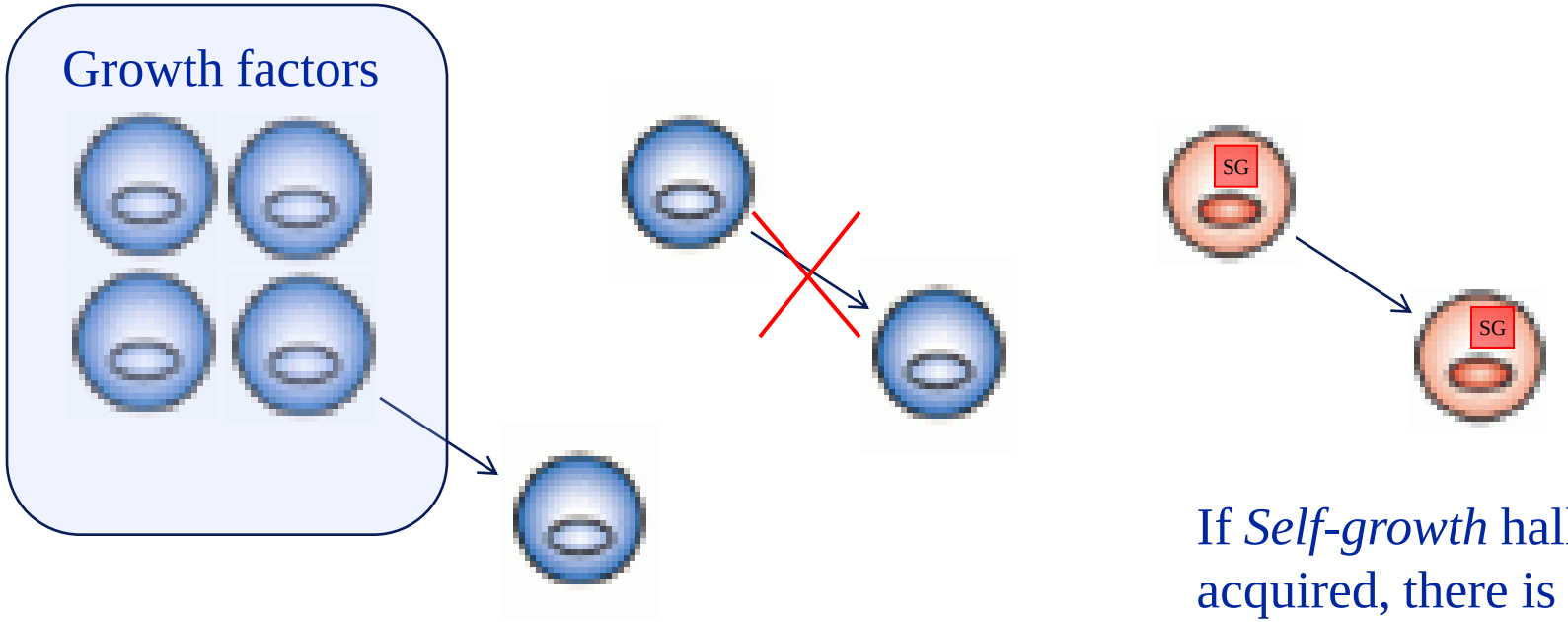
Each gene (hallmark) is mutated with a *1/m* chance of mutation

Daughter cell

Normal cell mitosis

1 iteration = 2.6 hours
5000 iterations = 75 weeks

AC Cell growth simulation. Self-growth hallmark (SG)

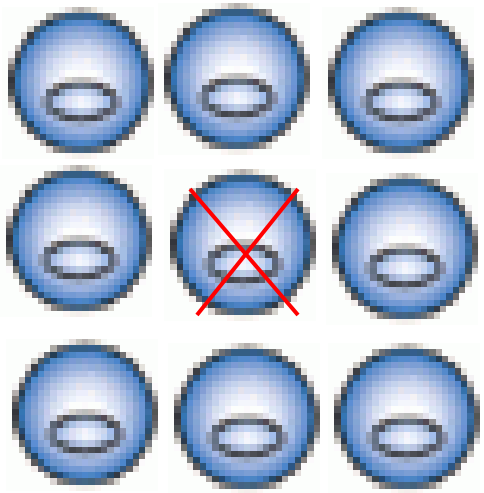


Normal cell mitosis

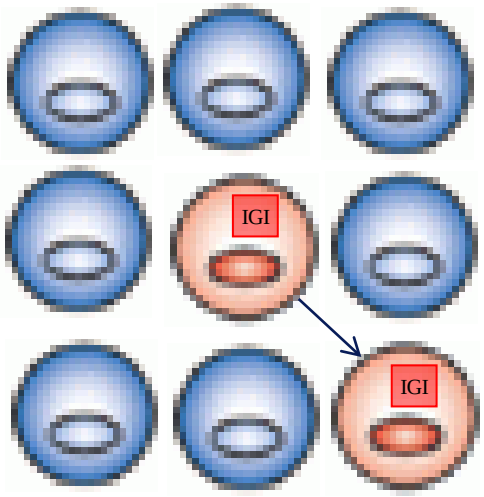
If *Self-growth* hallmark acquired, there is no need of growth factors

2D section of the grid

AC Cell growth simulation. Ignore growth inhibit hallmark (IGI)

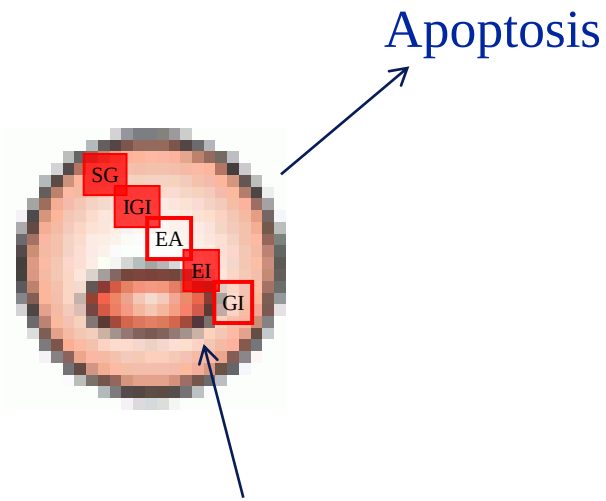


No empty space in the neighborhood

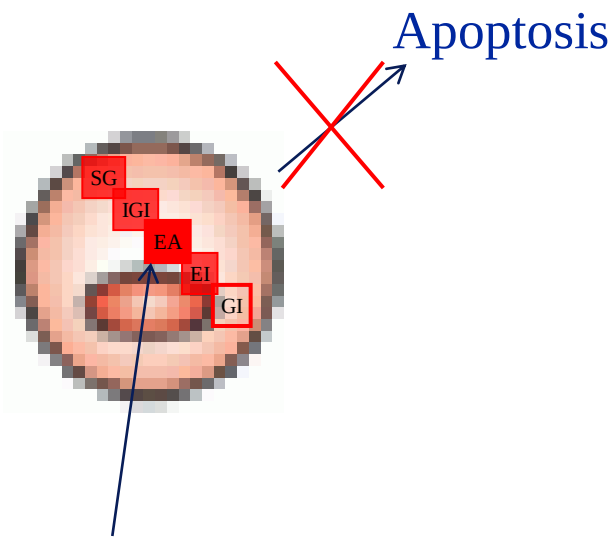


If *Ignore growth inhibit* hallmark acquired, a neighbor is killed (with probability *1/g*) to make room for mitosis

AC Cell growth simulation. Evade apoptosis hallmark (EA)

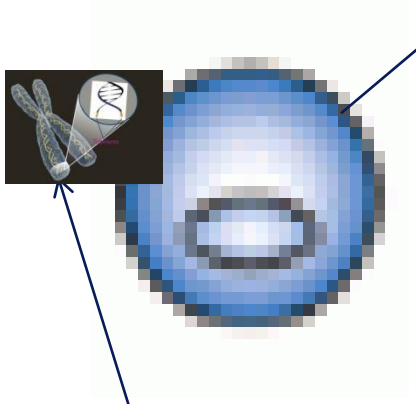


A cell with *n* hallmarks mutated has an extra *n/e* likelihood of dying each cell cycle.



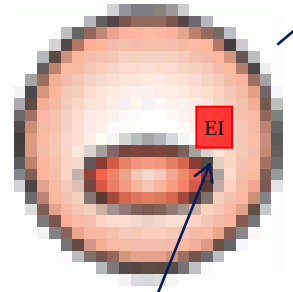
If Evade apoptosis hallmark (EA) is ON, there is not apoptosis

AC Cell growth simulation. Effective immortality hallmark (EI)



Cell death

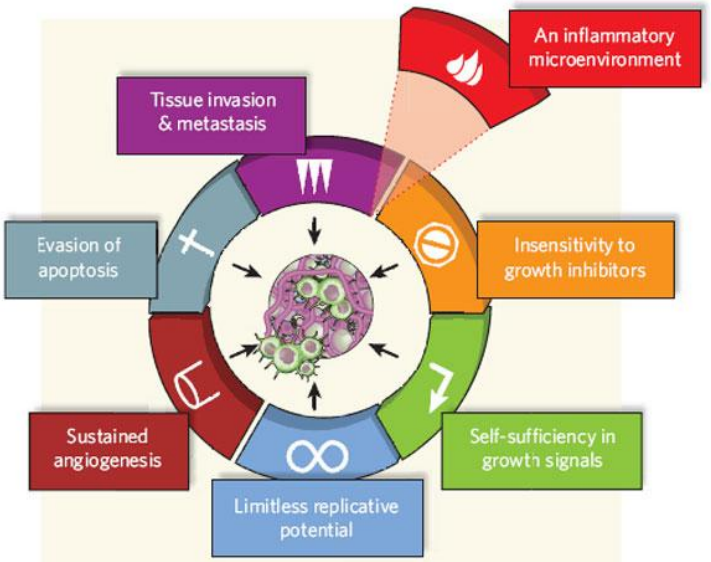
If telomere reaches 0, the cell dies.



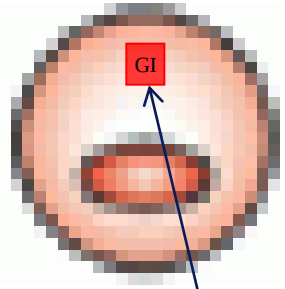
Can perform mitosis

Even if telomere length =0, if *Effective immortality* hallmark (EI) is ON, the cell can continue dividing.

AC Cell growth simulation. Genetic instability (GI)



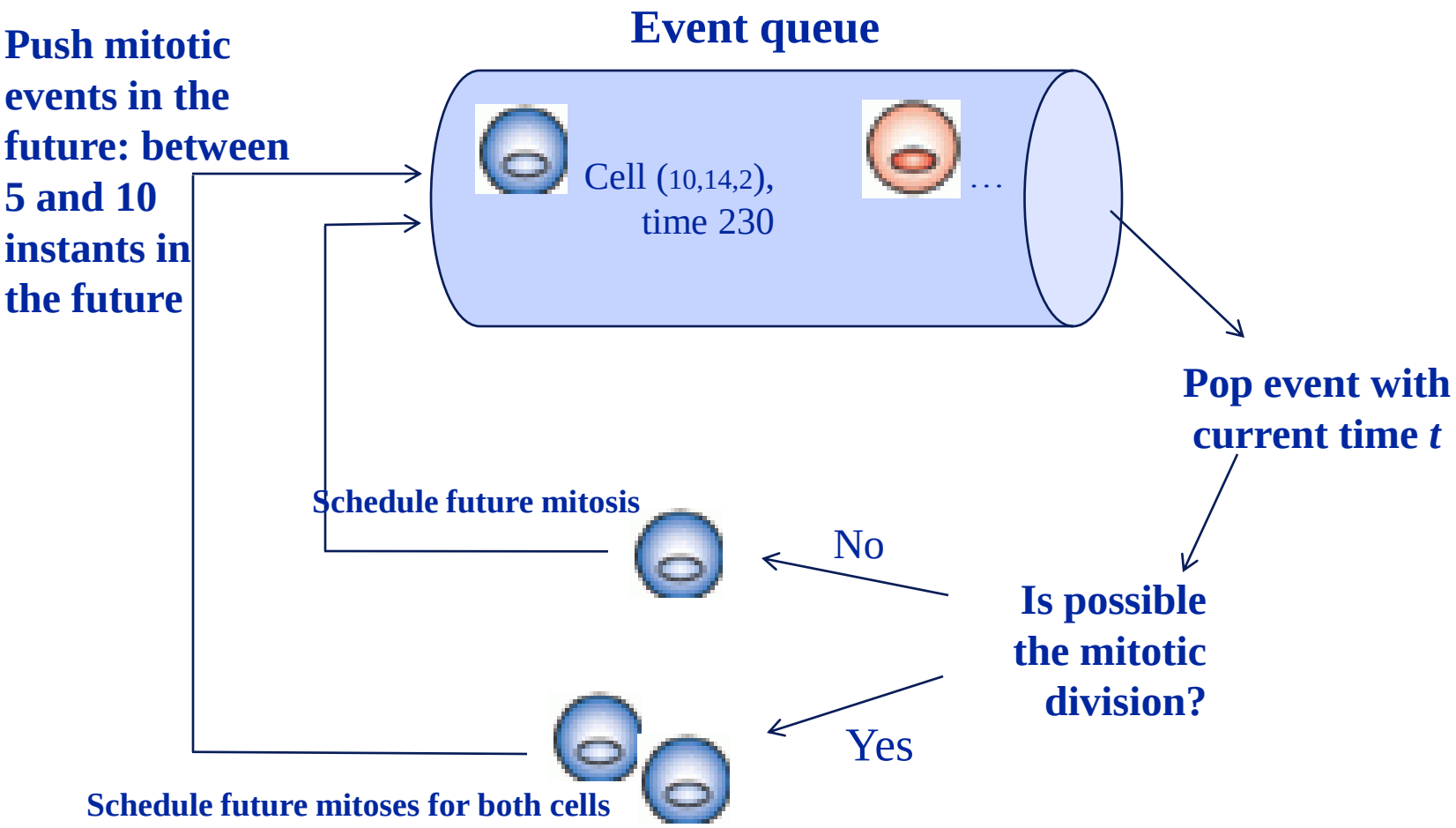
Hanahan and Weinberg (2011), "Hallmarks of Cancer: The Next Generation", *Cell*.



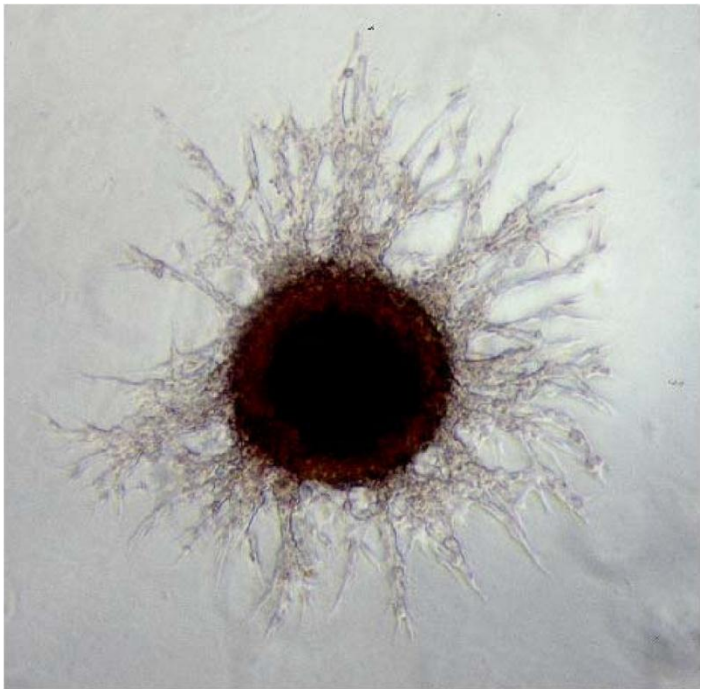
If *Genetic instability* factor (GI) is ON

Increment the cell base mutation rate by a factor (*i*)

Event model

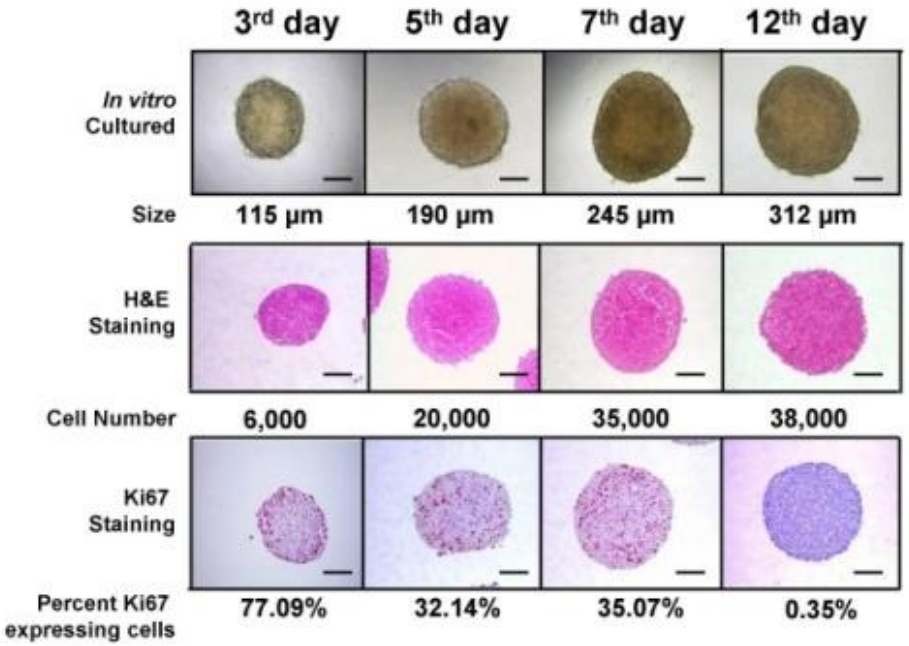

Event queue
Push mitotic events in the future: between 5 and 10 instants in the future
Cell (10,14,2), time 230
…
Pop event with current time *t*
Is possible the mitotic division?
No
Schedule future mitosis
Yes
Schedule future mitoses for both cells

Simulating cancer cell cultures: Multicellular spheroids



Microscopy image of a multicellular tumor spheroid, exhibiting an extensive branching system that rapidly expands into the surrounding extracellular matrix gel. These branches consist of multiple invasive cells.

Guiot *et al. Theoretical Biology and Medical Modelling* 2007 **4**:4 doi:10.1186/1742-4682-4-4

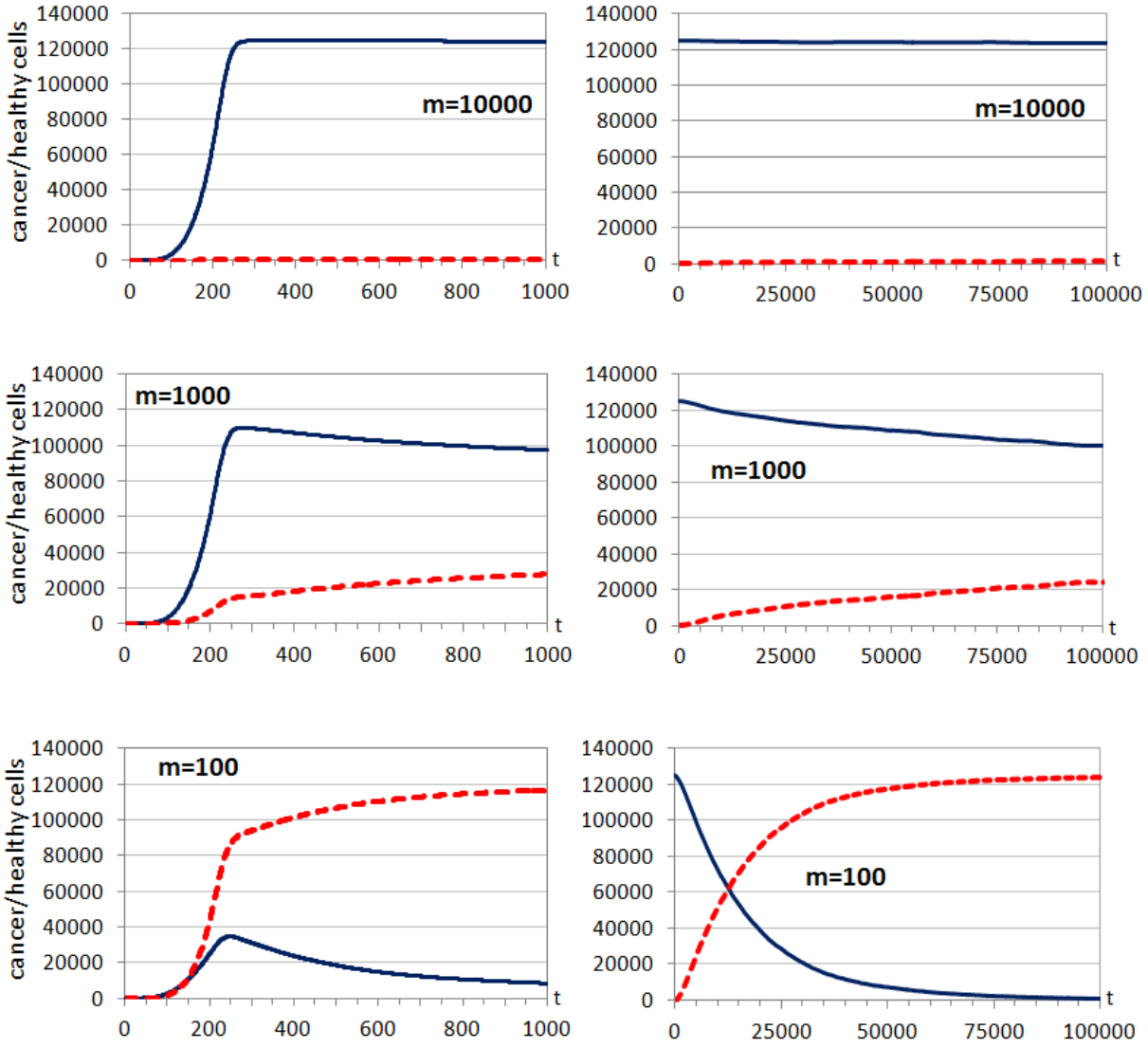


Growth stages of 3D-cultured CT26 colon cancer spheroid. Five hundred suspension cancer cells were dispensed into each well of a 48-well culture tray. Trays were then inverted and incubated during 12 days. Spheroids were collected on days 3, 5, 7 and 12 and processed for cell counting, spheroid diameter determination and immunohistochemical detection of Ki67-expressing cells. Scale bar: 100 µm.

Valcárcel M, et al.- J Transl Med (2008)

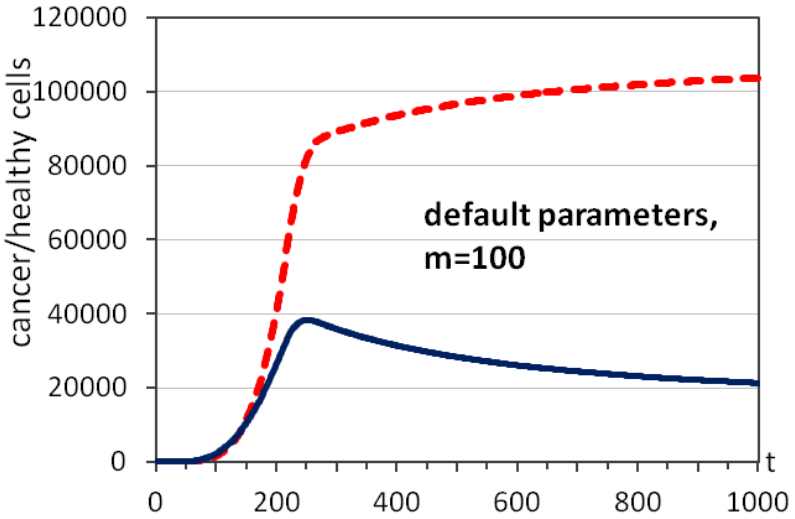
Some simulations and results

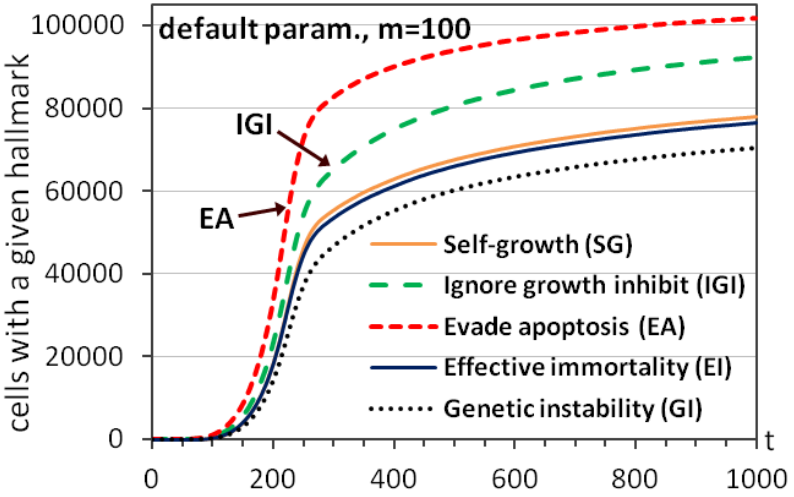
Dependence on hallmark parameters



Evolution through time iterations of the number of healthy cells (continuous lines) and cancer cells (dashed lines) for different base mutation rates (*1/m*)

Simulation run with default parameters and m=100

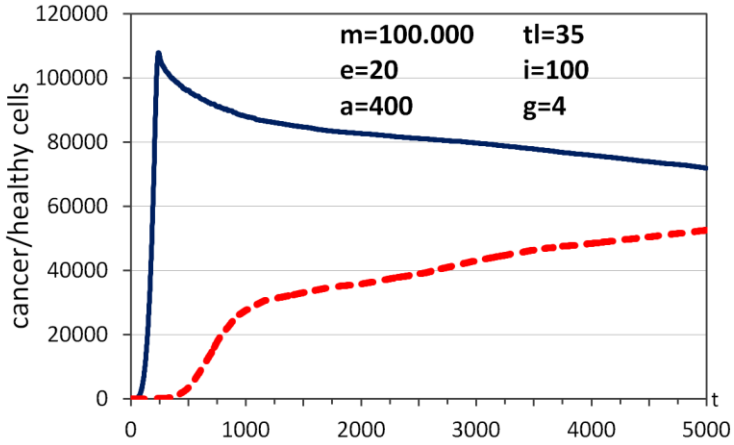


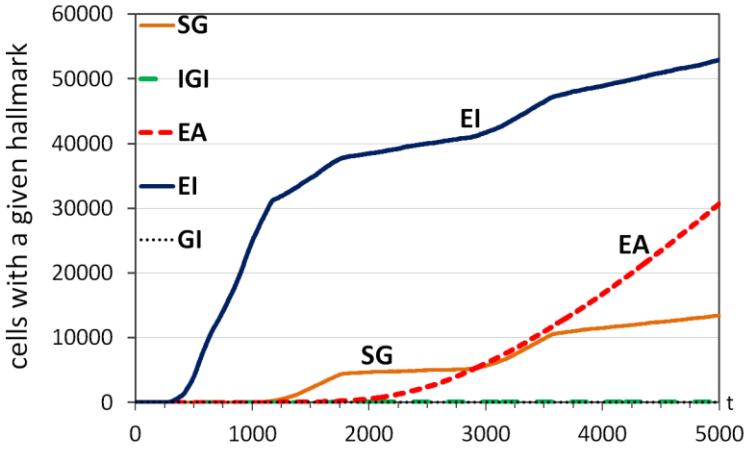


0

Grid size=125000, mutation rate=0.01

Simulation run with parameters that facilitates cancer appearance



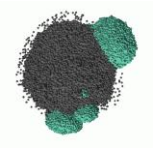


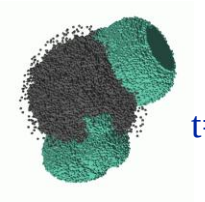
Grid size=125000, mutation rate=10^{-5}

Simulation run with parameters that facilitates cancer appearance


t=400


t=420


t=490

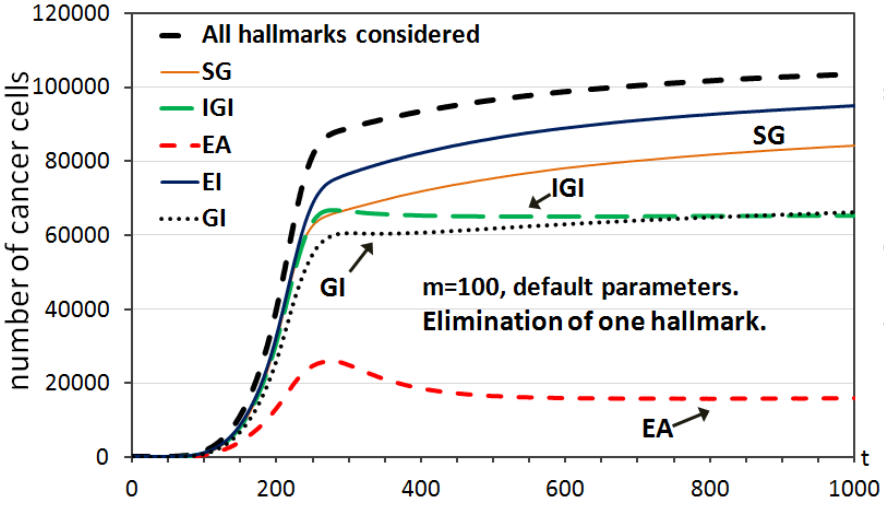

t=540

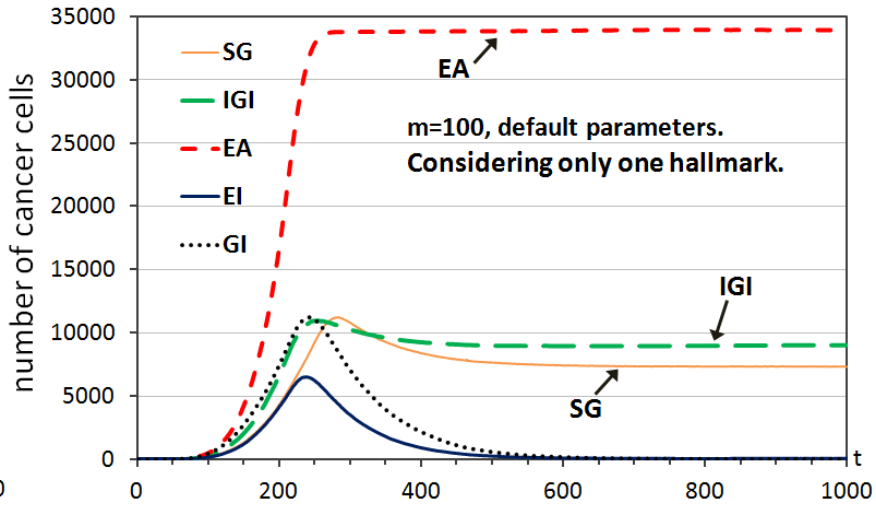
Snapshots of the cellular system at different time steps

Grid size=$10^{6,}$ mutation rate=10^{-5}

0

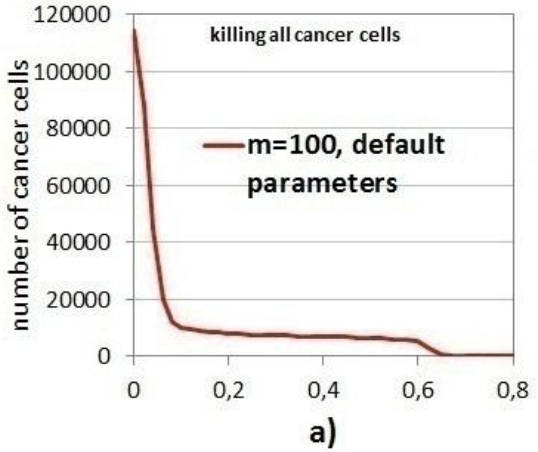
Relevance of hallmarks

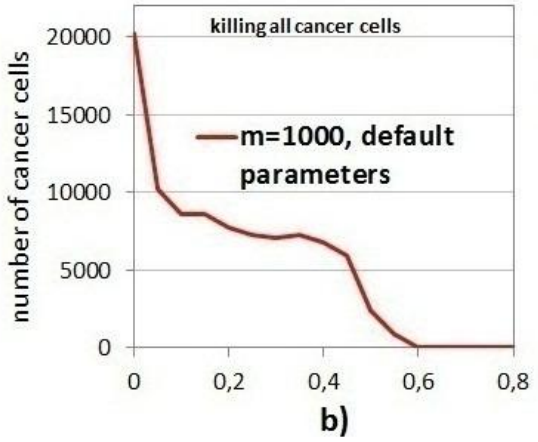


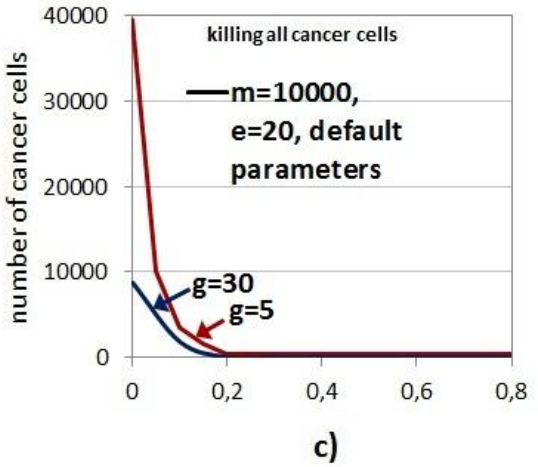


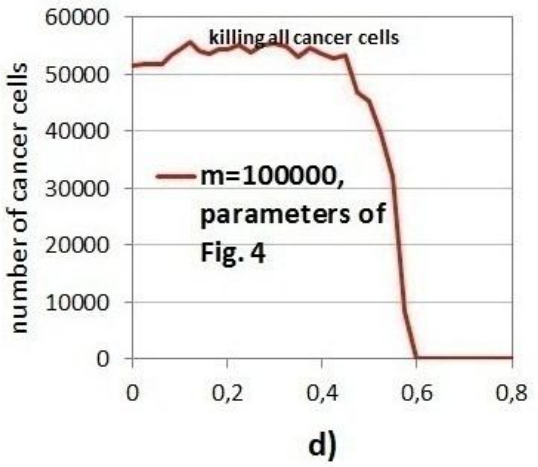
Effect of elimination of individual cancer hallmarks. Default parameters.

Behavior transitions

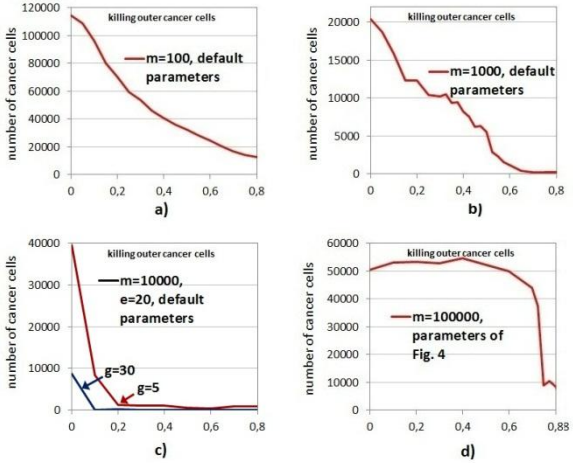






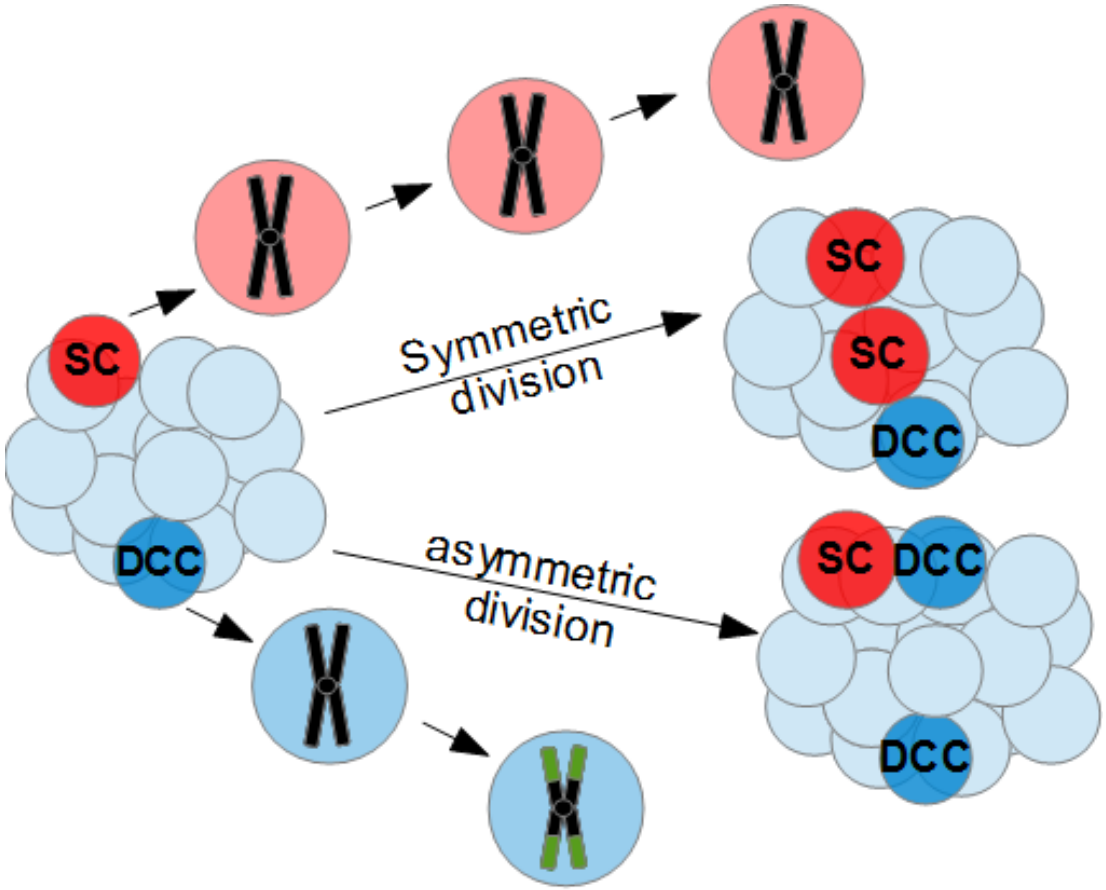


Killing only outer cancer cells



Effect of killing cancer cells during tumor growth for different killing probabilities and using four parameter sets.

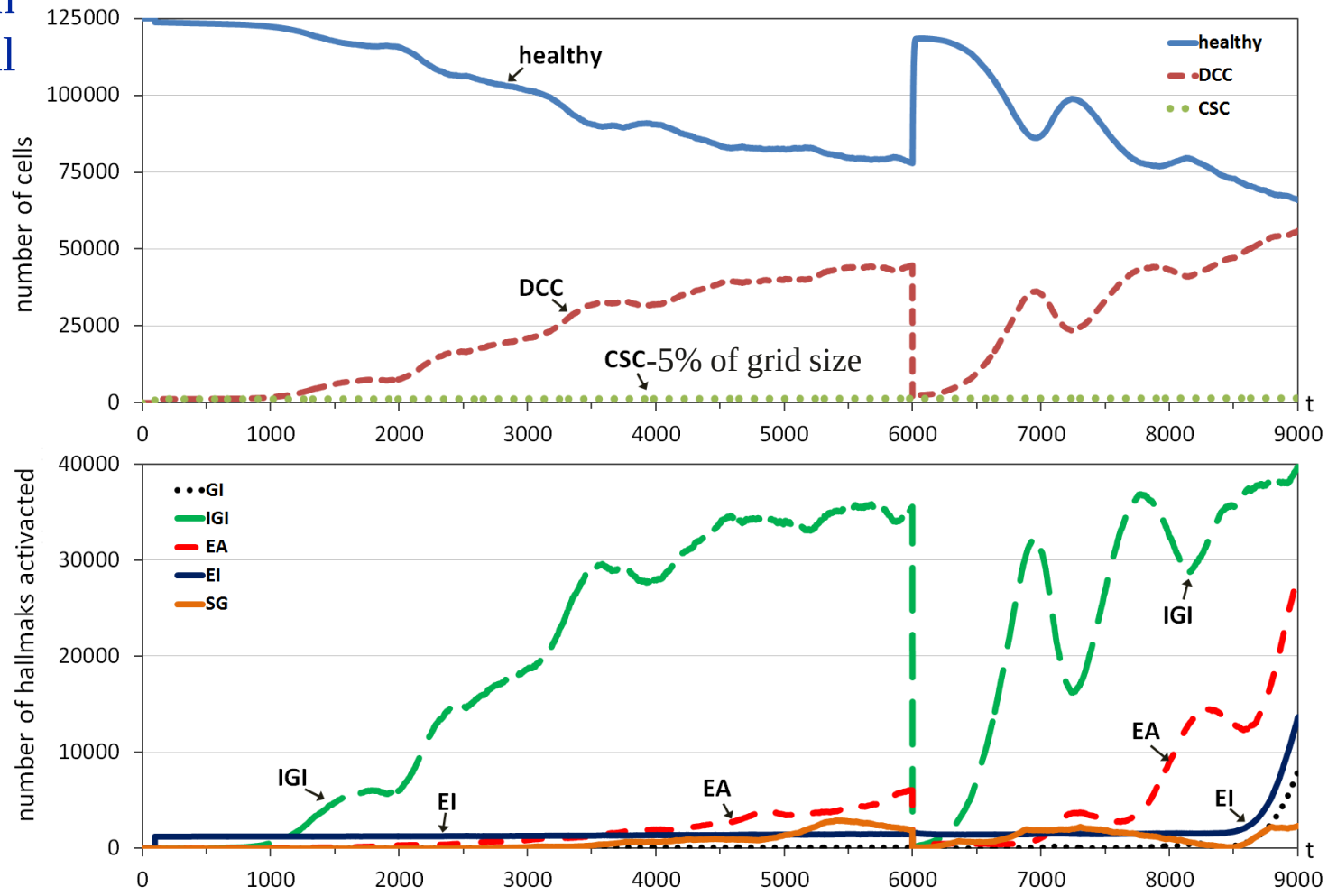
Cancer Stem Cells



CSCs can divide symmetrically or asymmetrically to produce Differentiated Cancer Cells (DCCs) with limited proliferative capability.

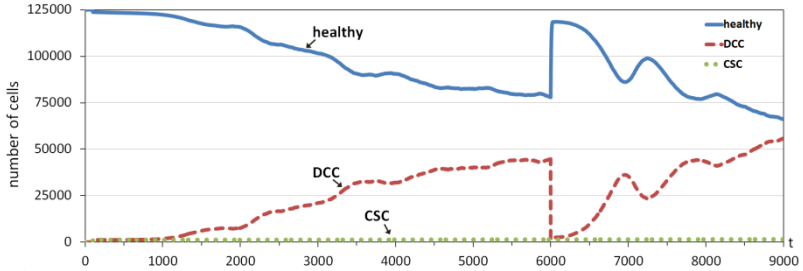
Simulating Cancer Stem Cells

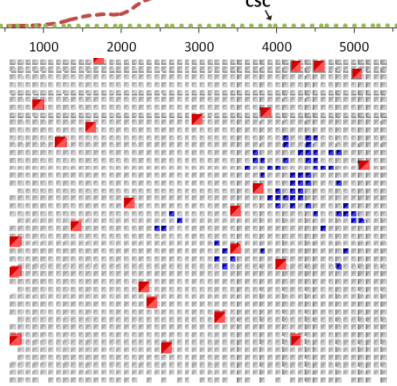
Example with high invasion potential



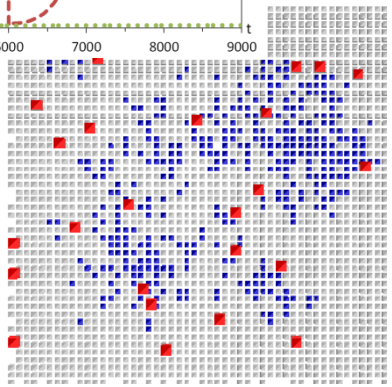
Introduction of Cancer Stem Cells (CSCs) in the multicellular system evolution. Standard parameters and g=5.

Simulating Cancer Stem Cells

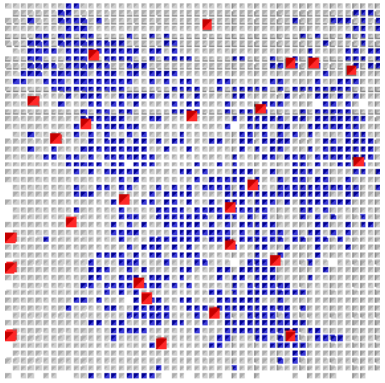




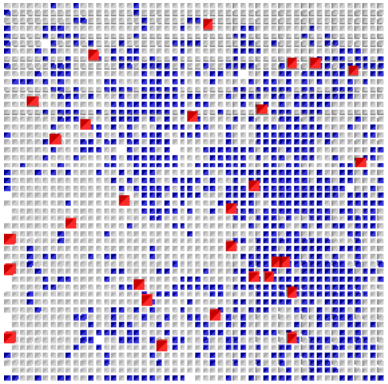
t=1500



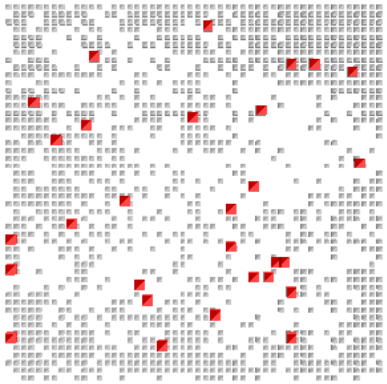
t=2500



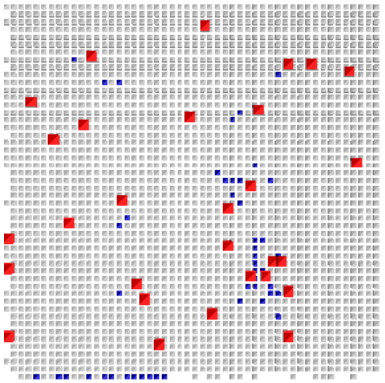
t=3500



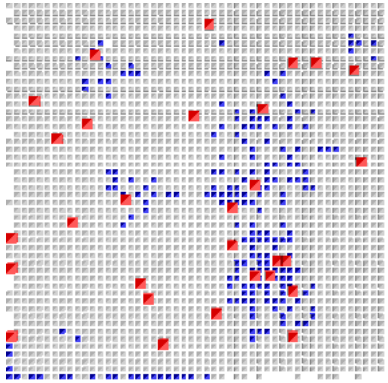
t=5500



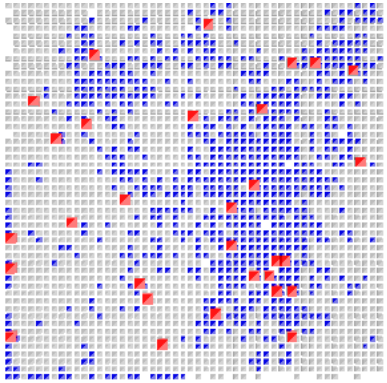
t=6000



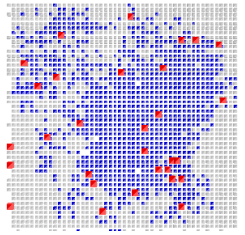
t=6250



t=6500



t=7000



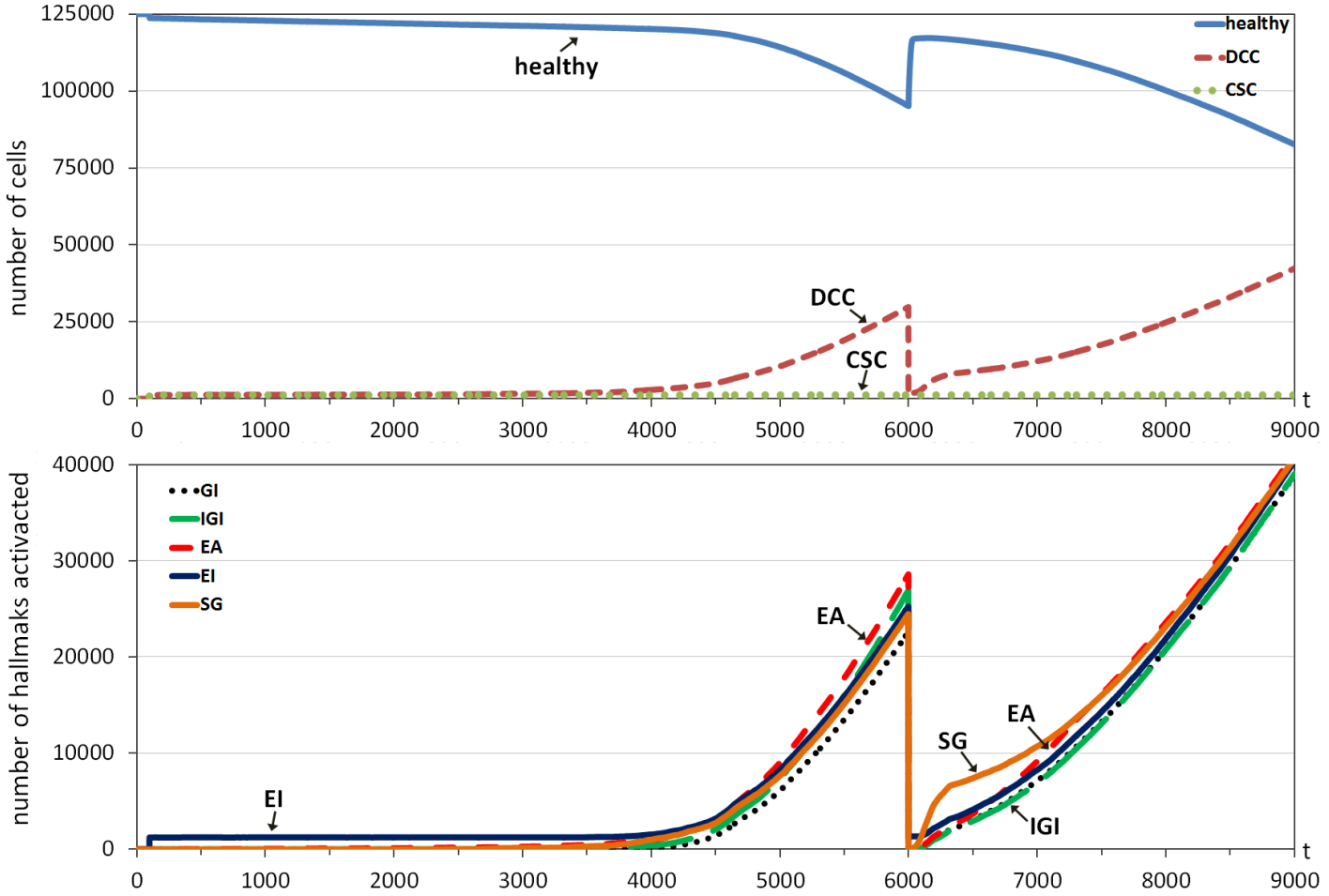
t=8500

Snapshots of 2D central sections

Gray: **Healthy** cells, Blue: **DCCs**, Red-enlarged size: **CSCs**.

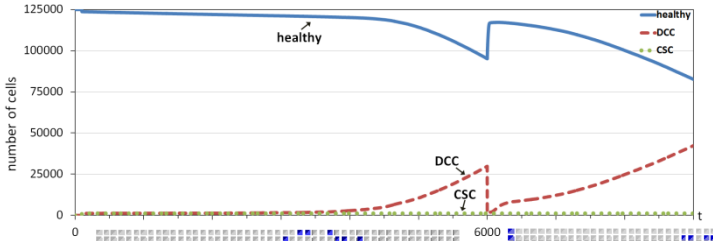
Simulating Cancer Stem Cells

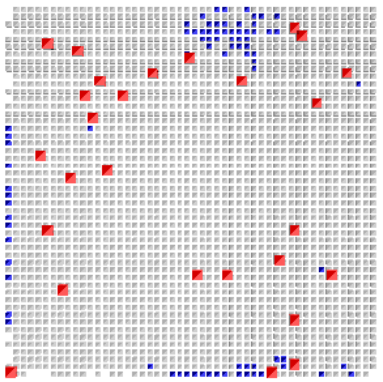
Example with high mutation rate



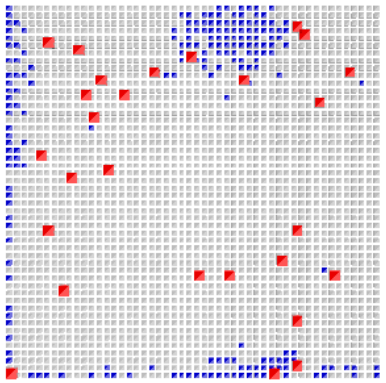
Standard parameters and m=1000.

Simulating Cancer Stem Cells

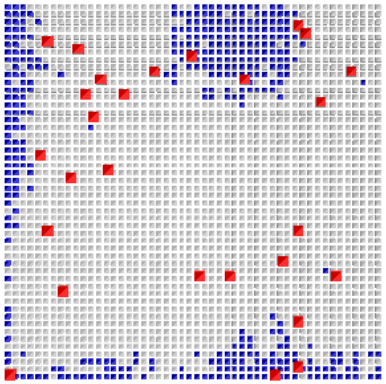




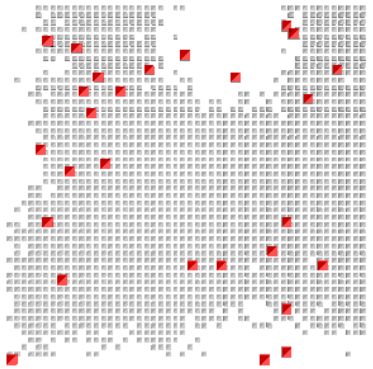
t=4500



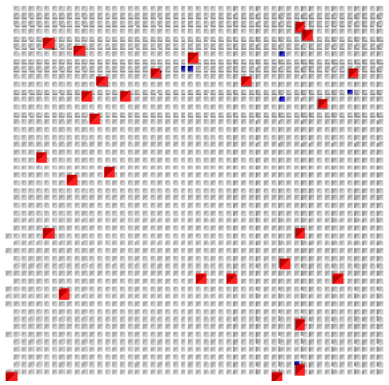
t=5000



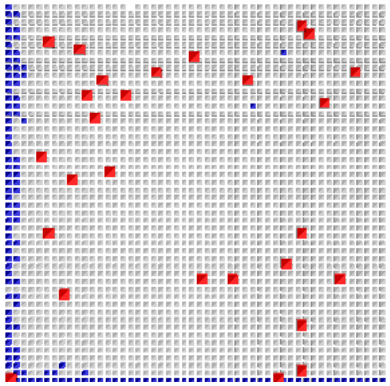
t=5500



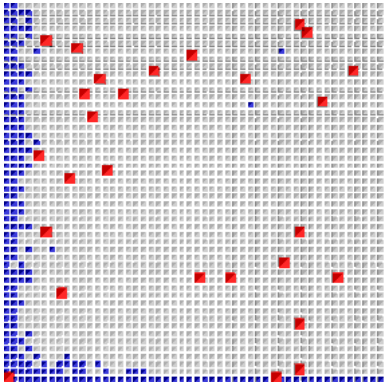
t=6000



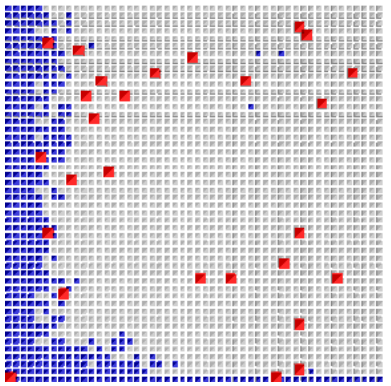
t=6100



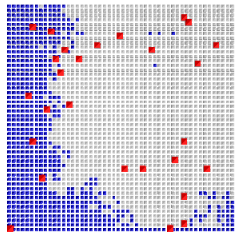
t=6500



t=7000



t=8000

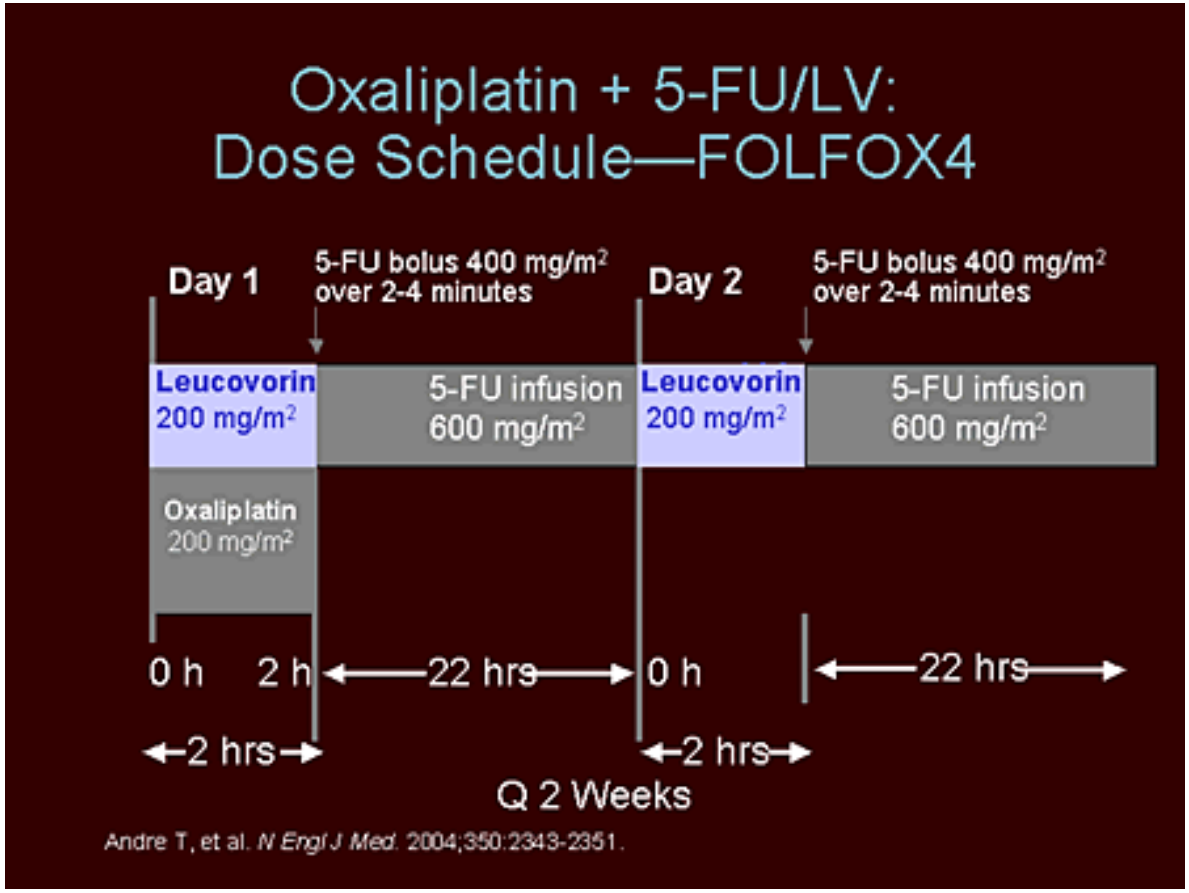


t=9000

Snapshots of 2D central sections

Gray: **Healthy** cells, Blue: **DCCs**, Red-enlarged size: **CSCs**.

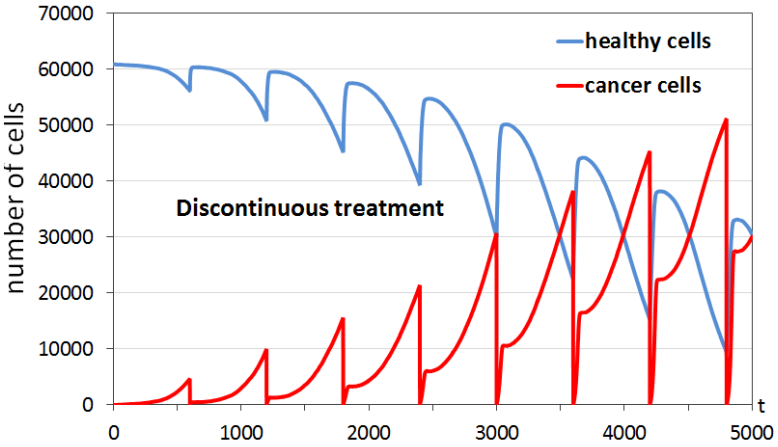
Treatment options

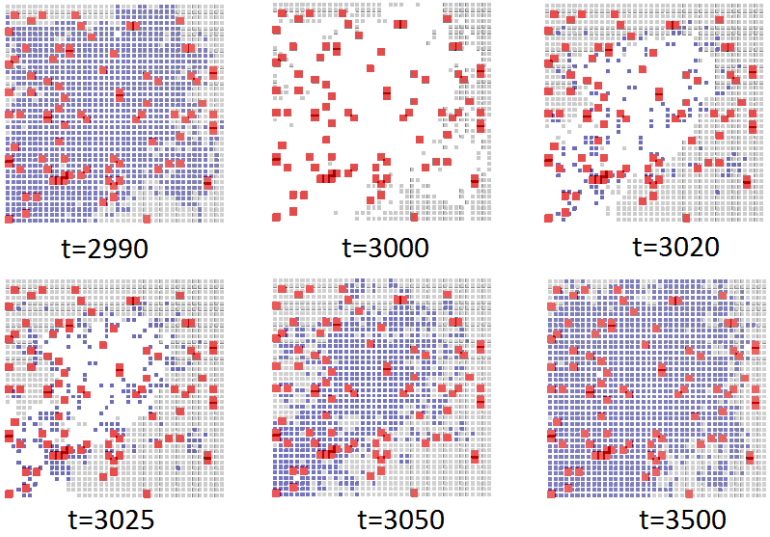


FOLFOX4 treatment for colon cancer

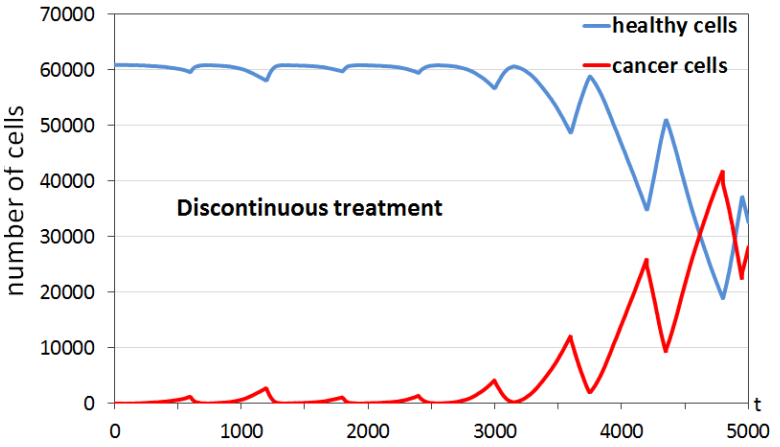
Treatment options

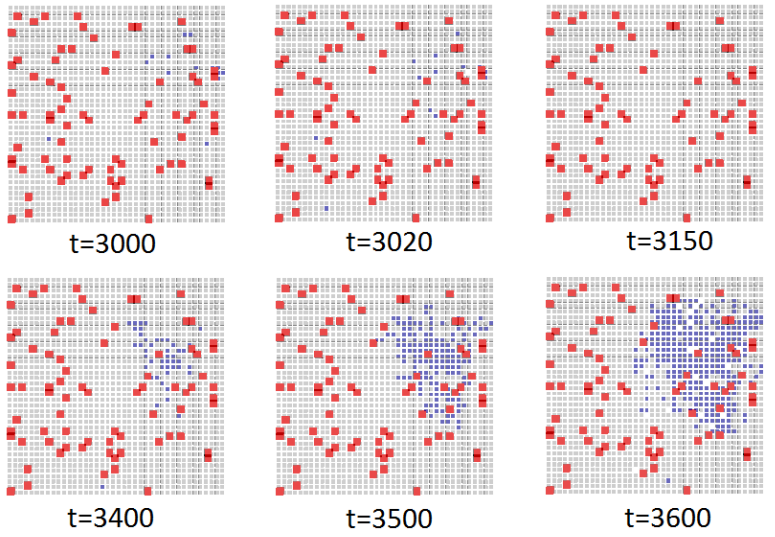
Treatment applied every 600 time iterations killing the 100% of the DCCs in such iterations.





Treatment applied every 600 time iterations killing the 5% of the DCCs during the next 150 iterations.

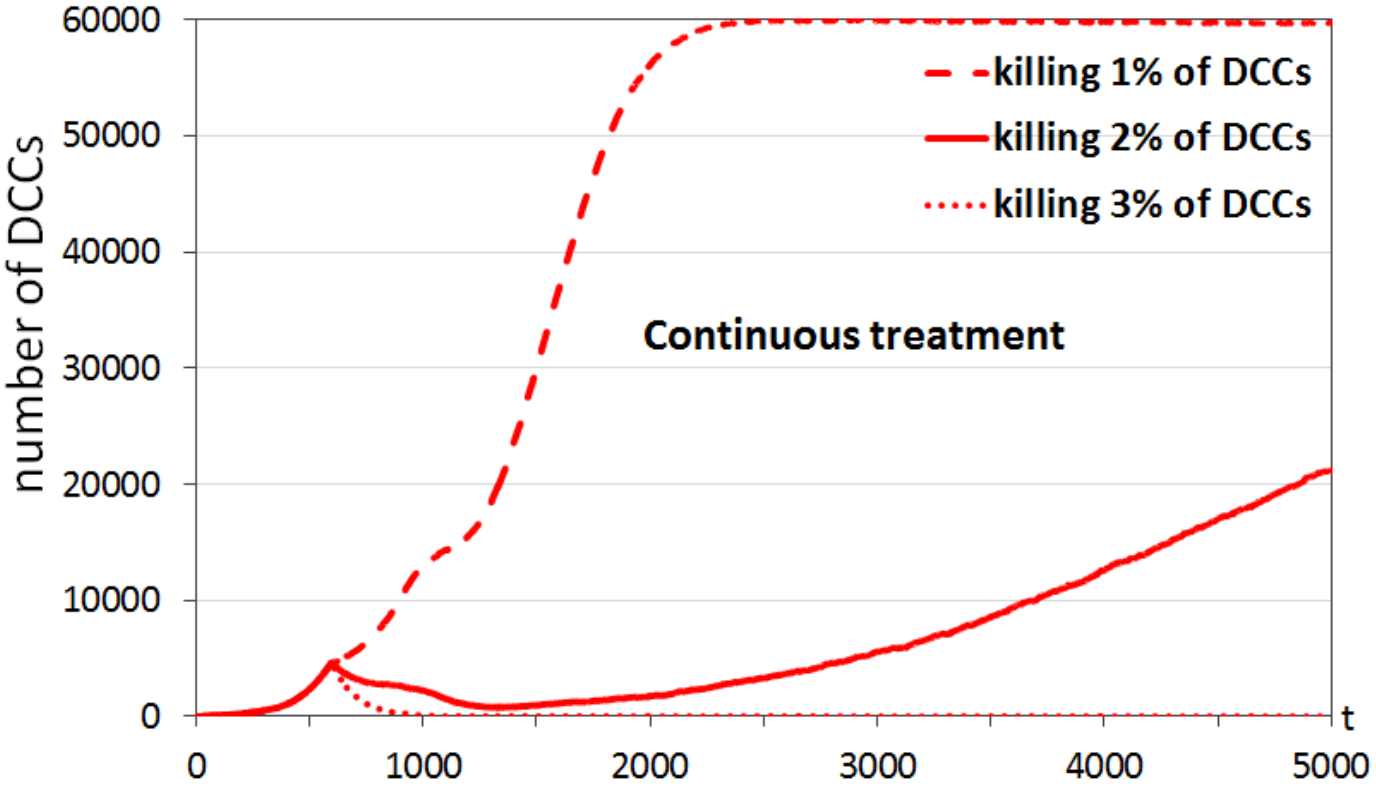




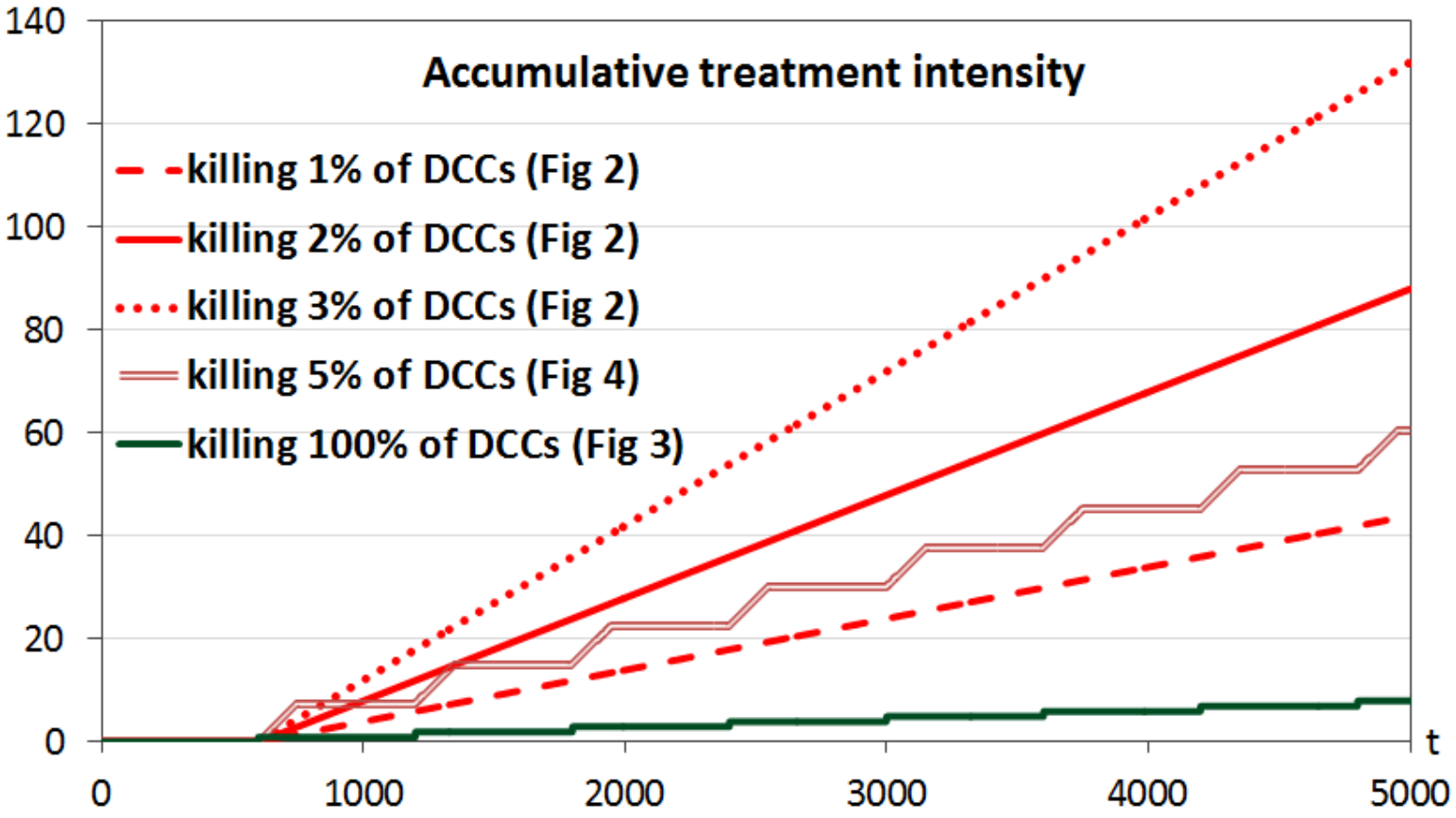
2D snapshots of the central part of the grid at given time iterations (Colors: Gray - healthy cells, Blue - DCCs, Red-enlarged size - CSCs).

Treatment options

Continuous treatments that kill 1%, 2% and 3% of DCCs are applied in every time iteration.



Treatment options



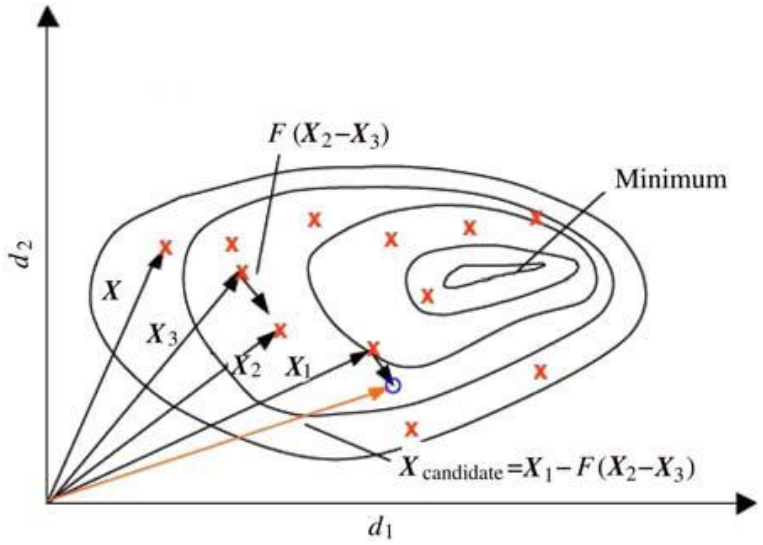
Accumulative treatment intensity across iterations

Differential Evolution

In this application: F:0.9

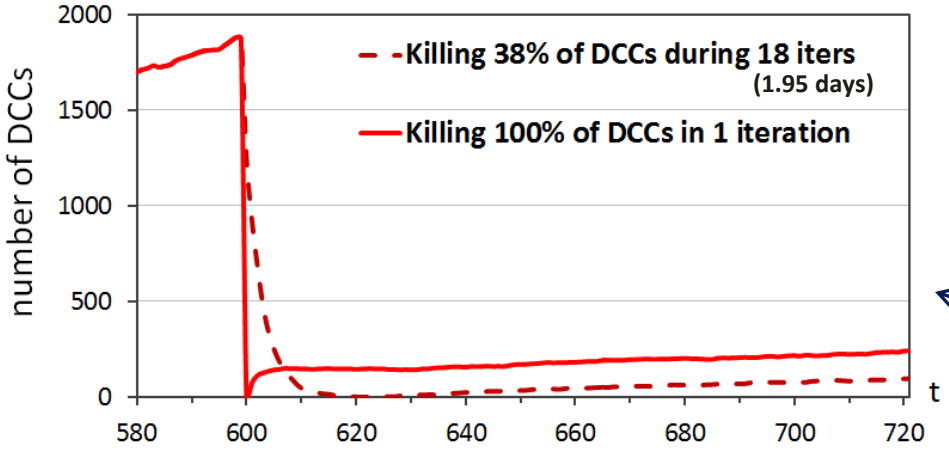
CR:0.9

X_1 selected with tournament

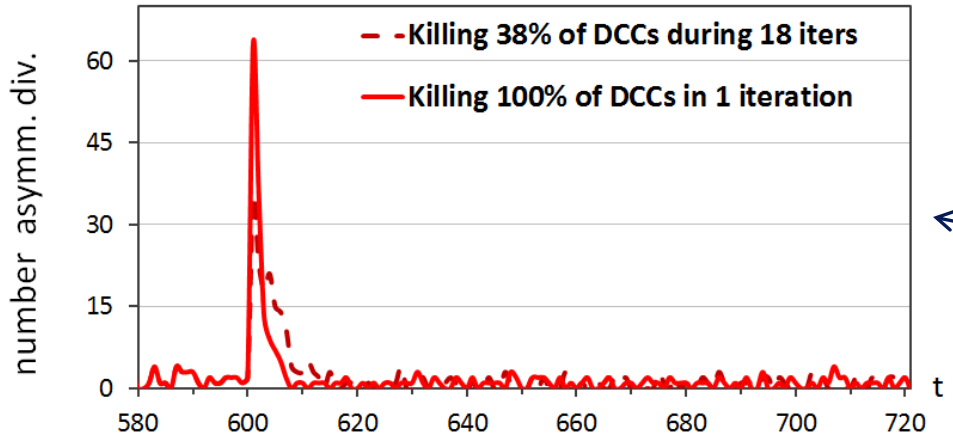


Encoded parameters: treatment intensity, duration and period.
50 generations, grid size of 64000 sites.
Hallmark parameters: g = 5, m = 1000, others - standard values
Fitness: DCCs after a given number of iterations (100) once the treatment is stopped +
Intensity used during the treatment application.

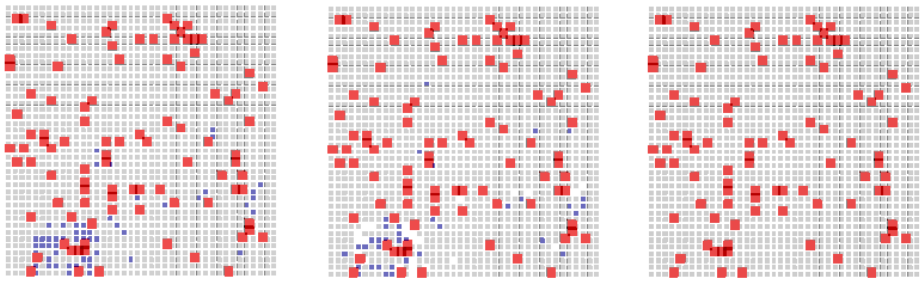
Evolved treatment strategies



Evolution of DCCs

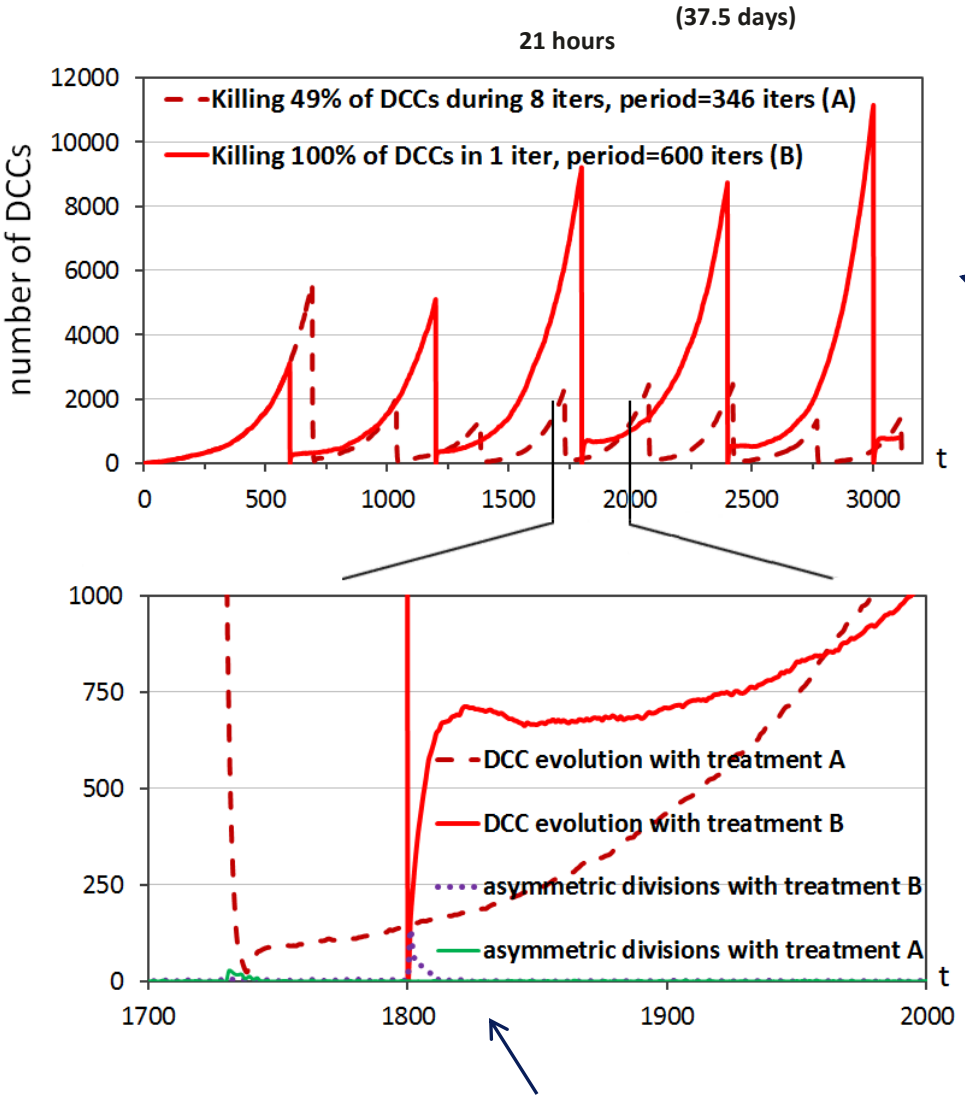


CSC asymmetric divisions



Treatment intensity and duration optimized. Treatments begin in time iteration t=600. The best evolved treatment kills 38% of DCCs in the next 18 time iterations. A high-intensity treatment that kills the 100% of DCCs only at t=600 is included for comparison.

Evolved treatment strategies



Evolution of DCCs

CSC asymmetric divisions

Treatment period, intensity and duration optimized. Treatments do not begin before time iteration t=600. The best evolved treatment kills 49% of DCCs during 8 time iterations and a period of 346 iterations (treatment A). A high-intensity treatment that kills the 100% of DCCs every 600 iterations is included for comparison (treatment B).

Conclusions

The hallmark implications for cell population dynamics are difficult to foresee without a simulation model.

Our focus was on the dependences of the first phases of cancer growth on the hallmark parameters:

- In a CSC context, the model predicts and is in agreement with the clinical observations describing **increased growth speed and enhanced invasion in the relapsing malignancy**.
- Using EC: Treatments should be maintained during very few days, avoiding high intensities and, consequently, with longer periods than the ones used in standard treatments. The aim is to **make more difficult CSC proliferation** and consequently their differentiation to minimize their future effect on a possible tumor regrowth

The modeling can lead to a better understanding and characterization of the underlying biological processes involved.

References

- Adami, C. (1998), *Introduction to artificial life*, Telos-Springer Verlag.
- Ilachinski, A. (2001), *Cellular automata. A discrete universe*. World Scientific.
- Langton, C.G. (1989), "Artificial life", *Artificial Life, SFI Studies in the Sciences of Complexity*, Vol. VI, C.G. Langton (Eds.), 1-47, Addison-Wesley.
- Monteagudo, A. and Santos, J. (2014), "Studying the capability of different cancer hallmarks to initiate tumor growth using a cellular automaton simulation. Application in a cancer stem cell context", *Biosystems* 115: 46-58.
- Monteagudo, A. and Santos, J. (2015), "Evolutionary optimization of cancer treatments in a cancer stem cell context", *Proceedings Genetic and Evolutionary Computation Conference - GECCO 2015*, 233-240.
- Monteagudo, A. and Santos, J. (2015), "Treatment analysis in a cancer stem cell context using a tumor growth model based on cellular automata", *Plos One*, doi: 10.1371/journal.pone.0132306.
- Santos, J. (2007), *Vida Artificial. Realizaciones Computacionales*, Servicio de Publicaciones de la Universidad de A Coruña.
- Santos. J. and Diéguez, M. (2011), "Differential evolution for protein structure prediction using the HP model*", Inter. Work-Conf. on the Interplay between Natural and Artificial Comp., LNCS* 6686:323-333.
- Santos, J. and Monteagudo, A. (2014), "Analysis of behaviour transitions in tumour growth using a cellular automaton simulation", *IET Systems Biology*, doi: 10.1049/iet-syb.2014.0015.
- Santos, J., Villot, P., Diéguez, M. (2014), "Emergent protein folding modeled with evolved neural cellular automata using the 3D HP model", *Journal of Computational Biology* 21(11):823-845.

http://www.dc.fi.udc.es/ai/~santos/