Algorithms of Life

François Fages

Lifeware group http://lifeware.inria.fr/
on Computational Systems Biology and Optimization

National Institute of Research in Computer Science and Control

Inria Saclay, Palaiseau, France

Model Building: Two Contradictory Perspectives

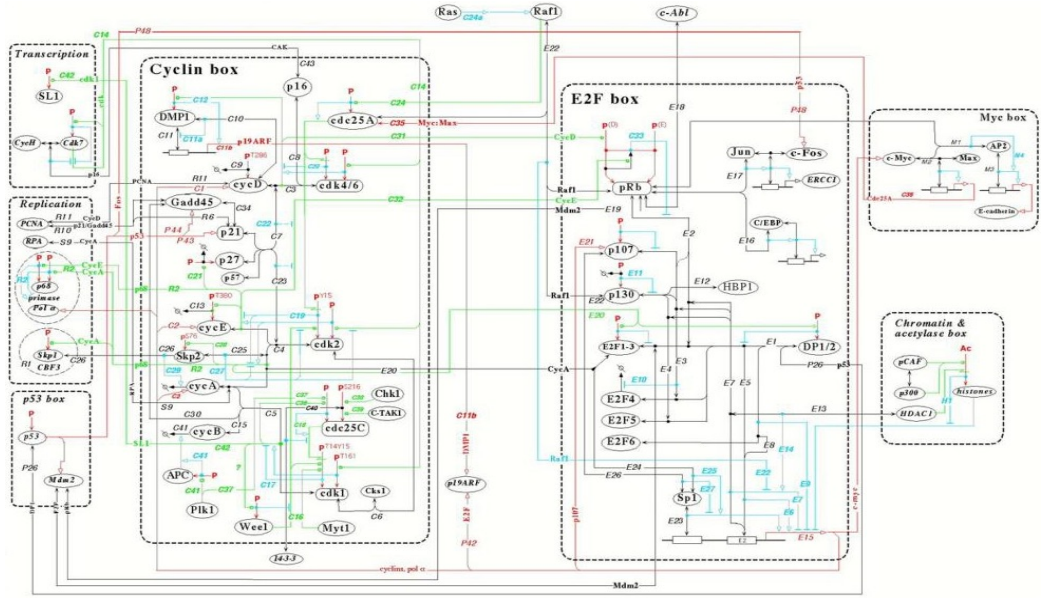
1) Model for representing knowledge: the more detailed the better (do not miss any known information)

2) Model for answering a concrete question: the more abstract the better (get rid of irrelevant information)

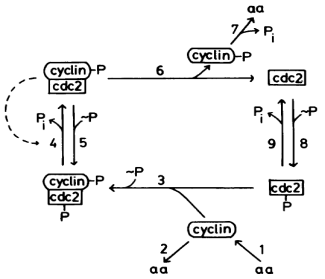
The Digital Twin approach needs reconciling those two perspectives for a large class of questions

Subcellular level: chemical reaction/influence networks (CRN, SBML exchange format)

1) Generic annotated interaction graph model



2) Reduced graph models



2) Reduced boolean models

2) Reduced ODE models

$$
\begin{aligned}
d[C2]/dt &= k_6[M] - k_8[\sim P][C2] + k_9[CP] \\
d[CP]/dt &= -k_3[CP][Y] + k_8[\sim P][C2] - k_9[CP] \\
d[pM]/dt &= k_3[CP][Y] - [pM]F([M]) + k_5[\sim P][M] \\
d[M]/dt &= [pM]F([M]) - k_5[\sim P][M] - k_6[M] \\
d[Y]/dt &= k_1[aa] - k_2[Y] - k_3[CP][Y] \\
d[YP]/dt &= k_6[M] - k_7[YP]
\end{aligned}
$$

2) Reduced stochastic CTMC models (intrinsic noise)

Requires general notions of model structure reductions and model dynamics abstractions

Model Structure Reductions by Subgraph Epimorphisms [Gay F- Martinez Soliman Solnon *DAM* 2014]

Purely graphical notion of model reduction by 4 graph operations (compatible with slow-fast ODE reductions)

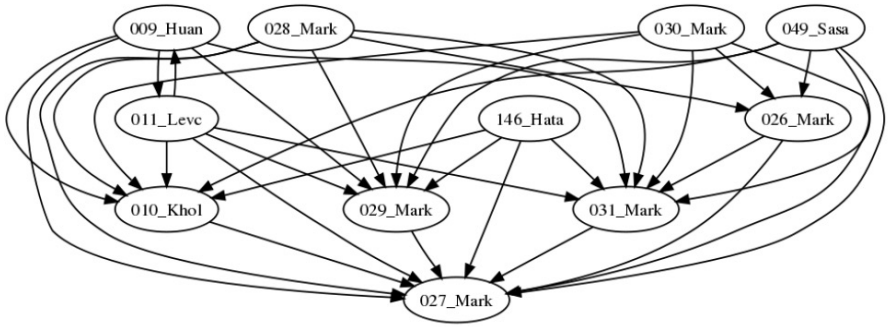
- Deletion of (irrelevant) species
- Deletion of (neglectable) reactions or influences
- Merging of (similar) species
- Merging of (aggregable) reactions or influences

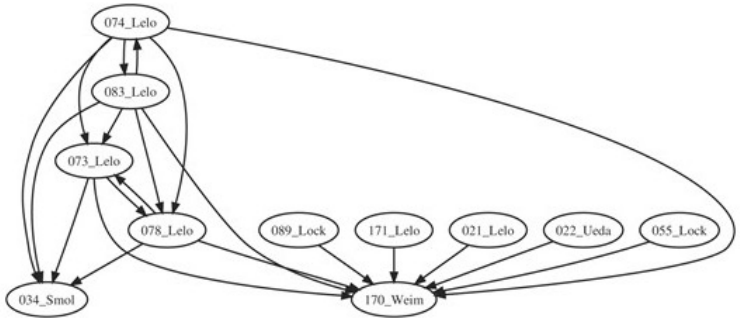
SISO: subgraph isomorphism

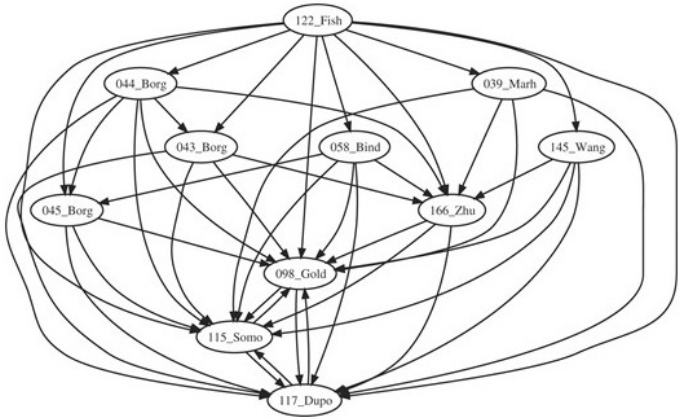
EPI: epimorphism

SEPI

Automatic reconstruction of model hierarchies in BioModels [Gay Soliman F- *Bioinformatics* 2010]







Possible with graphical CRN models, not with their ODE instanciation [Inferring CRNs from ODEs. F- Gay Soliman. *TCS* 2015]

Model Dynamics Abstractions

Theory of Abstract Interpretation for computer programming [Cousot Cousot *POPL* 1977]

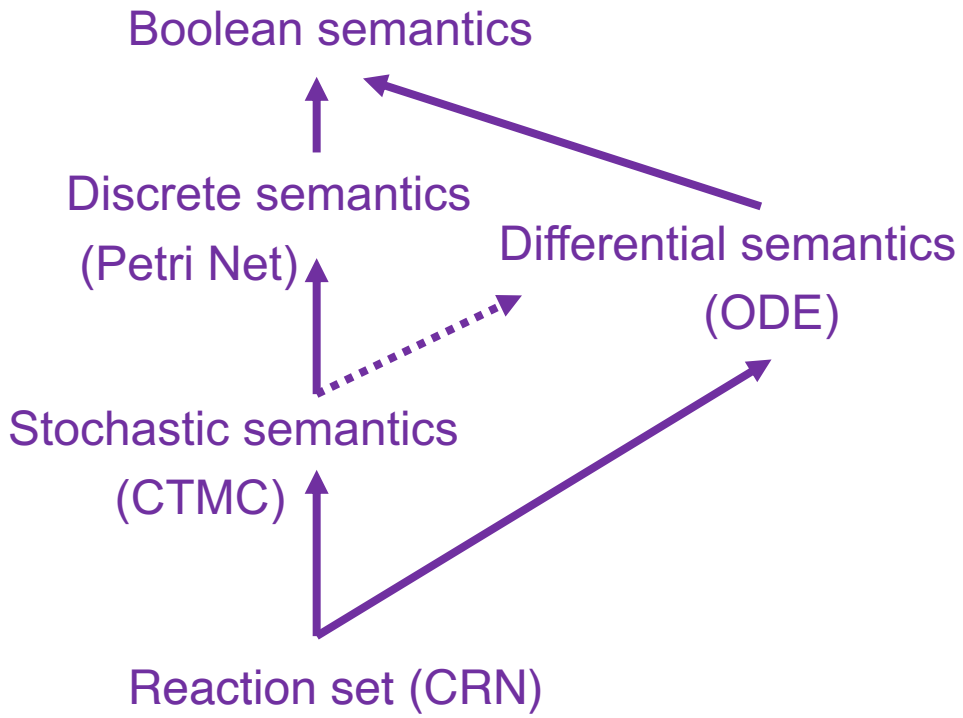
Biochemical Reaction Networks CRN as a programming language

Theorem (abstract interpretation →) Galois connections between the syntactical, stochastic CTMC, Petri net and Boolean trace semantics [F- Soliman *TCS* 2008]

If a behavior is not possible in the Boolean semantics (verifiable by model-checking) it is not possible in the stochastic semantics for any reaction rates.

Boolean model behaviors may correspond to rare events.

Theorem (approximation ⇢) For large numbers of molecules the ODE behavior approximates the stochastic behaviors' mean [Kurtz 1978, 1992]



Analog Computation with CRN Programs

Theorem (Turing-completeness of finite CRNs with ODE semantics) [F- Le Guludec Bournez Pouly CMSB 2017]

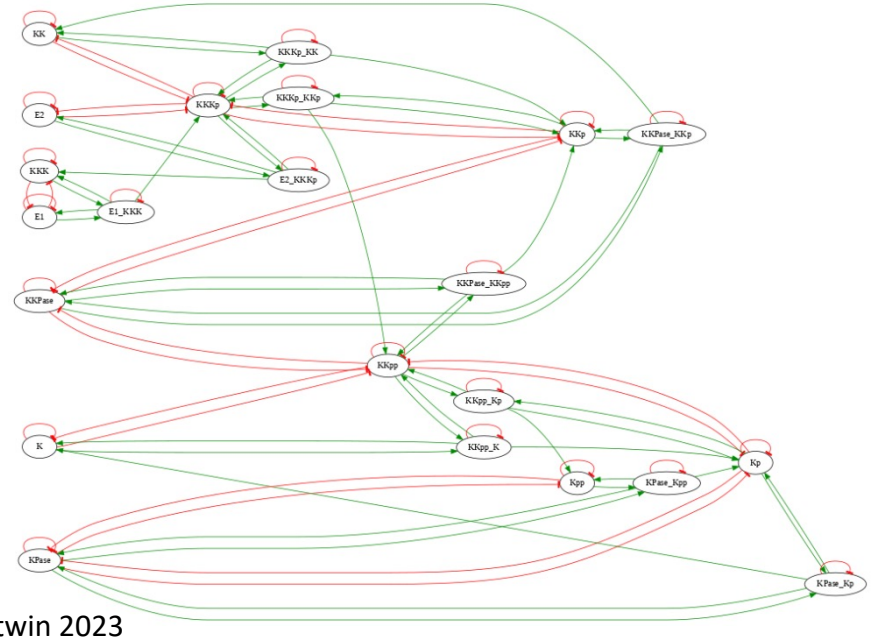
Any computable real function (i.e. by a Turing machine with arbitrary requested precision given in input) can be computed by a finite CRN with mass action law kinetics and at most bimolecular reactions.

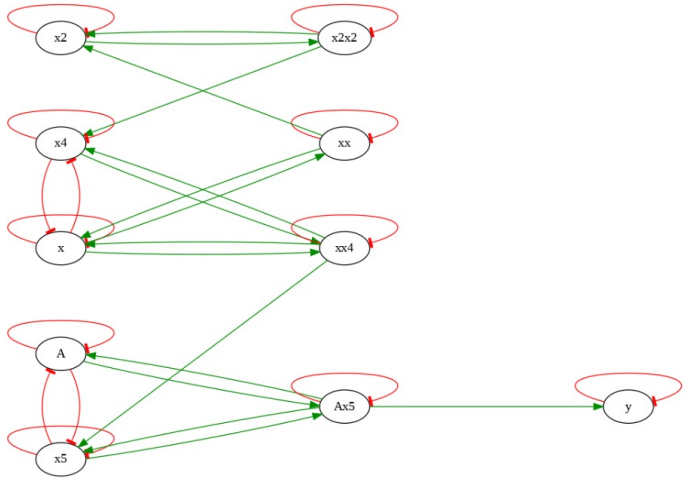
Theorem (Online computation, robust stabilization) [Hemery F- CMSB 2022]

The set of real functions computable online by a CRN is precisely the set of real algebraic functions.

Example Natural MAPK CRN structure Biocham-compiler-generated CRN for same I/O function $\frac{x^5}{k+x^5}$

Stiff sigmoid analog-digital converter function of MAPK signaling





Cell Population Level: Hard-coded Multiscale Interfaces

Yeast single cell gene expression control to achieve a given wave objective

Cell Hog pathway ODE model

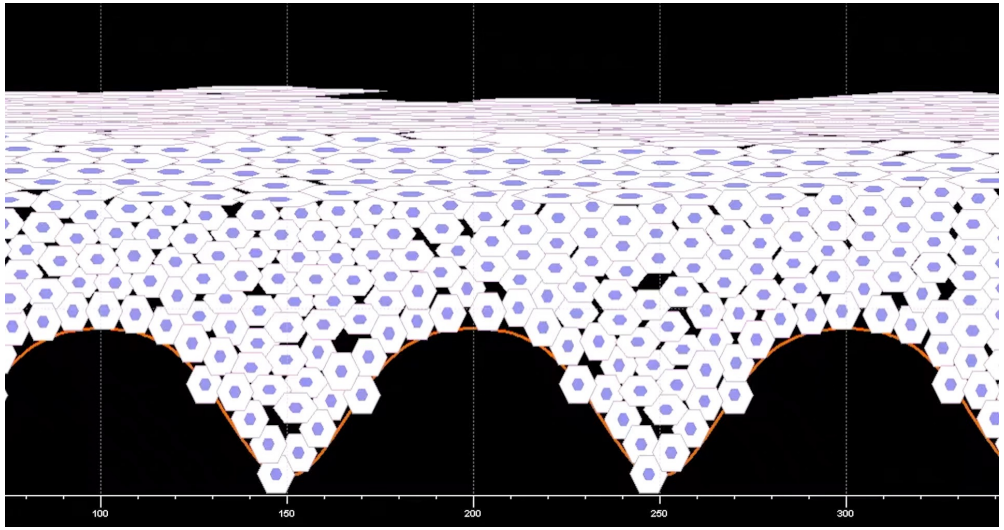
reduced to just one delay τ

in ODE model-based control

Jannis Uhlendorf, Gregory Batt, Pascal Hersen et al. Long-term model predictive control of gene expression at the population and single-cell levels. *PNAS* 109(35) 2012.

Multicellular Tissue Level: Hard-coded Model Plugins

- Influential factors on model atopic dermatitis ? pH, microbiome, model parameters
- Qualitative comparison to skin barrier thickness measurements



1. Skin cell CRN model sensitive to pH elevation

2. Commensal and pathogenic bacteria population model
   - Quasi-stability at the time scale of the experiments (2 days)
   - Followed by a population switch at longer time scale (e.g. 10 days)

directly used through SBML interface in a

3. Multi-agent skin tissue model (EPISIM)
   - No bacterial population switch on long time scale due to tissue dynamics

E. Greugny, J. Bensaci, F- , G. Stamatas.
Computational modeling predicts impaired barrier function and higher sensitivity to skin inflammation following pH elevation.
Journal of Experimental Dermatology, 2022.

Eléa Thibault Greugny, Georgios N Stamatas, F- Stability versus Meta-stability in a Skin Microbiome Model.
CMSB 2022.

Algorithmic Models

Population of N lymphocytes
- with N_A pathogens/antigens
- with affinity score S

Successively infected by
- pathogens $a(t)$
- during $\tau(t)$
- with malignity cost $C(t)$

Clonal expansion of best scored
Eradication of other lymphocytes
Regularization of over expression

Algorithm 1: Clonal Proliferation Algorithm

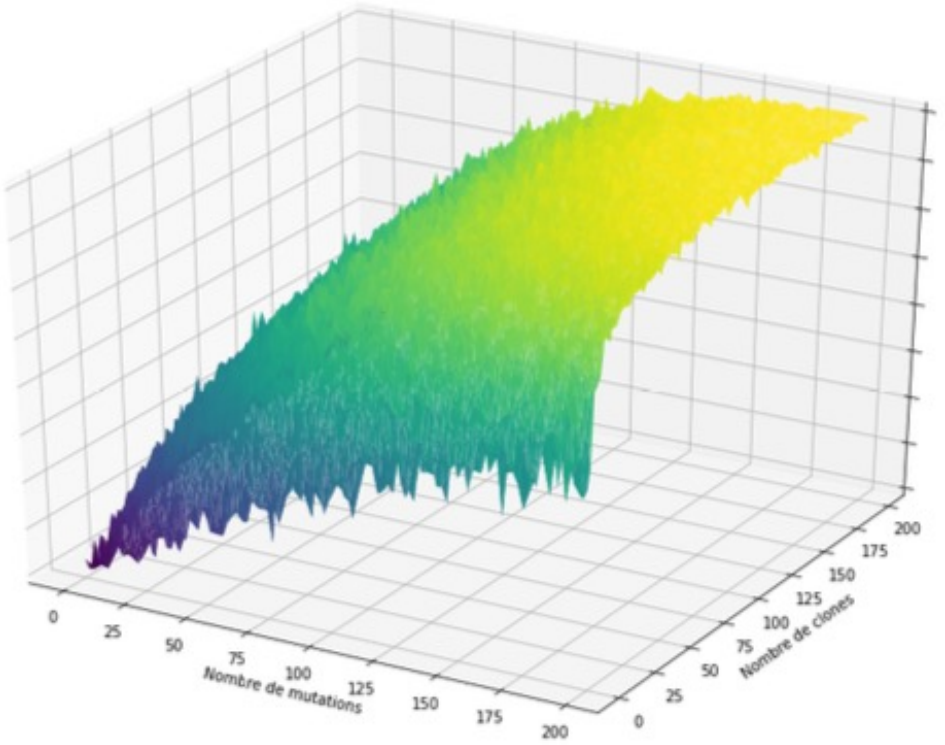
Result: Costs C, infection durations τ
Data: $N, N_A, t_{max}, T, S(\cdot)$
Generation of the naive repertoire $\mathcal{R}_L = \{L_1, \ldots, L_N\}$;
Generation of an antigen sequence $\mathcal{A} = \left\{\left(a_1, \mu_{a_1}\right), \ldots, \left(a_{N_A}, \mu_{a_{N_A}}\right)\right\}$;
Initialization of the cost $C = [\,]$ and time list $\tau = [\,]$;
for $t < t_{max}$ **do**
    Introduction of (a_t, μ_{a_t}) randomly picked in $\mathcal{A}$ in $\mathcal{R}_L$;
    Reaching of a score $S(\mu_{a_t})$ with a time $\tau_t^{a_t}$ with notably a sub-optimal L_t^* ;
    $C_t \leftarrow C(\mu_{a_t}, \tau_t^{a_t})$;
    $\tau_t \leftarrow \tau_t^{a_t}$;
    Expansion of some clones of L_t^* in $\mathcal{R}_L$. Eradication of other lymphocytes.;
    if *t is a multiple of the period T* **then**
        Regularization of $\mathcal{R}_L$;
    end if
end for
return C, τ

Immune Repertoire Construction

Algorithm 3: Construction

Result: $\mathcal{R}_M$
Data: d, N, N_A

Generation of the naive repertoire $\mathcal{R}_L = \{L_1, \ldots, L_N\}$;
Generation of an antigen sequence $\mathcal{A} = \{a_1, \ldots, a_{N_A}\}$;
foreach $a_j \in \mathcal{A}$ **do**
  $f_j \leftarrow [f_{ij}]_{i \in [1,N]}$; //
  $kL \leftarrow \mathbf{kArgMax}_{i \in [1,N]}(f_j)$;
  foreach $L_i \in kL$ **do**
    $L_i^* \leftarrow \mathbf{BestMutant}\left(\mathbf{Mutation}\left(\mathbf{Clones}(L_i, f_{ij})\right)\right)$
  end foreach
  $f_j^* \leftarrow [f_{ij}^*]_{\{L^* \in kL^*\} \cup \{L \in \mathcal{R}_L - kL\}}$;
  $\mathcal{R}_M \leftarrow \mathbf{UpdateMemory}(\mathcal{R}_M, f_j^*)$;
end foreach
return $\mathcal{R}_M$



(a) Affinity of mutants as a function of N_{clones} and $N_{mutations}$

Immune Repertoire Evolution
for naive L, memory M and plasma P cells

Algorithm 4: Evolution

Result: $\mathcal{R}_L, \mathcal{R}_M, \mathcal{R}_P$
Data: $d, N_L, N_P, N_M, t_{max}$
Input: $\mathcal{A}$

Generation of $\mathcal{R}_L = \{L_1, \ldots, L_{N_L}\}$ and $\mathcal{R}_P = \{P_1, \ldots, P_{N_P}\}$;
$\mathcal{R}_M \leftarrow$ **Construction**$(\mathcal{R}_L, \mathcal{A})$;
for $t \leftarrow 0$ **to** t_{max} **do**
  Pick $a_t \in \mathcal{A}$;
  $\{L_t^i\}_{i \in \mathcal{I}} \leftarrow$ **RecoAntigen**(a_t) ;
  foreach $i \in \mathcal{I}$ **do**
    $L_t^i, M_t^i, P_t^i \leftarrow$ **Metadynamic**$(L_{t-k_i}^i, M_{t-k_i}^i, P_{t-k_i}^i)$;
    $\tilde{P}^i{}_t \leftarrow$ **Mutation**(P_t^i);
  end foreach
  Update $\mathcal{R}_L, \mathcal{R}_M, \mathcal{R}_P$;
  Random mutation of $\tilde{L} \leftarrow$ **RandLymph**$(\mathcal{R}_L)$;
end for
return $\mathcal{R}_L, \mathcal{R}_M, \mathcal{R}_P$

L,M,P population dynamics equations

$$\begin{cases} L_t = (L_{t-k} - C)\, B^k + C\, e^{-k\delta} \\ P_t = \gamma^k P_{t-k} + \dfrac{\alpha}{2}\left[(L_{t-k} - C)\,\dfrac{B^k - \gamma^k}{B - \gamma} + C\dfrac{e^{-k\delta} - \gamma^k}{e^{-\delta} - \gamma}\right] \\ M_t = \zeta^k M_{t-k} + \dfrac{\alpha}{2}\left[(L_{t-k} - C)\,\dfrac{B^k - \zeta^k}{B - \zeta} + C\dfrac{e^{-k\delta} - \zeta^k}{e^{-\delta} - \zeta}\right] \end{cases}$$

Mutations of Antigens

Mutation of antigens as bit inversion

QR code representation of
Antigens and Lymphocytes Boolean vectors
of size an integer square

Algorithm 2: Mutation of an antigen a

Result: A
Input: $a, N_m, \mathcal{L} \equiv \{L \mid L\mathcal{R}a\}$

$A = [\,]$;
for $k \in [1, N_m]$ **do**
  foreach $e \in epitopes\,(a)$ **do**
    $\hat{f} \leftarrow \max_{L \in \mathcal{L}} f_{e,L}$;
    $P \leftarrow$ **ProbaMutation**$(\hat{f})$;
    With probability P, change epitope e of antigen a;
  end foreach
  $A \leftarrow a$;
end for
return A

Pattern Training and Recognition

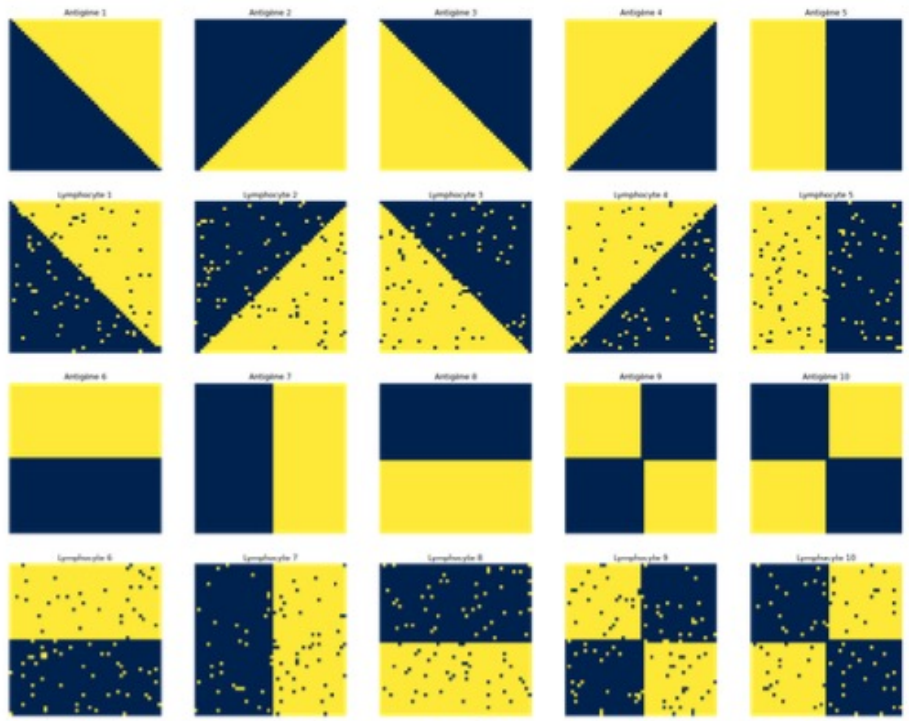


Figure 5.5: Representations of **ten lymphocytes and the respective antigen** they most closely approximate after 2 000 iterations of Algorithm `Evolution`.

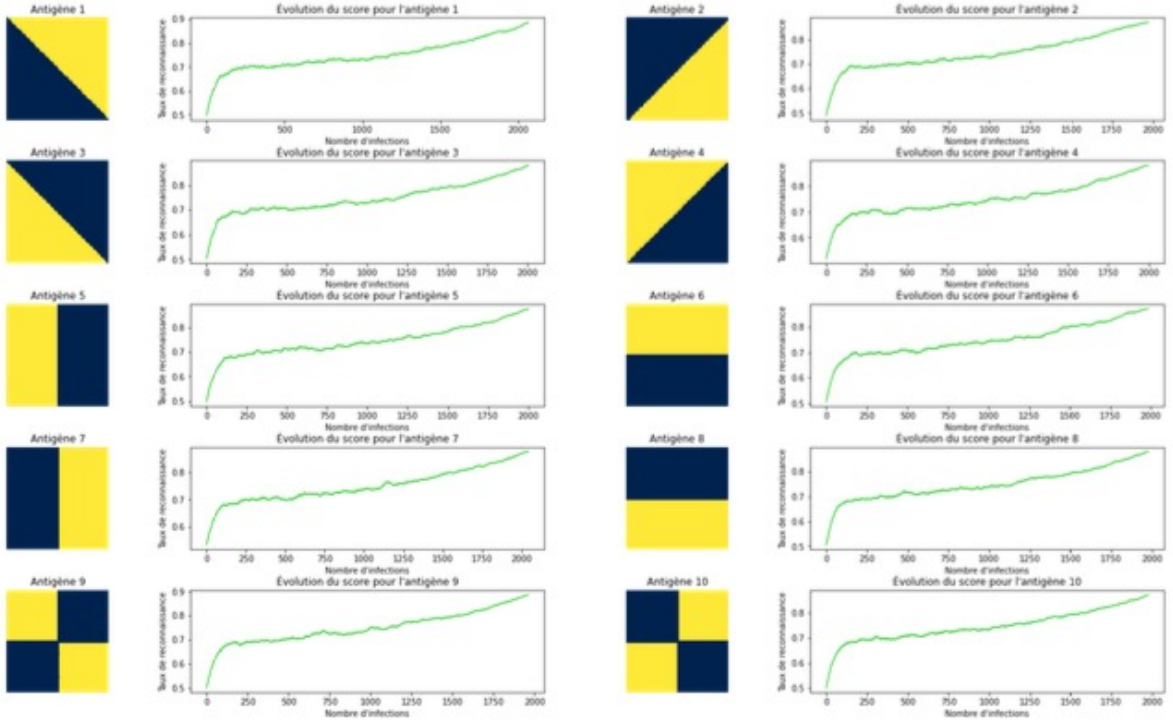


Figure 5.6: **Temporal growth of the affinity score** for the ten lymphocytes each approximating a target antigen over a horizon of 2 000 iterations.

Play with Vaccination Scenarios

Vaccine injection develops population of specific (memory and plasma) cells

- which promptly respond to infection by nearby pathogen
- with the risk of creating an overly targeted immune repertoire.

1st injection	Booster freq.	Comment
None	None/Low	No significant effect
None/Low	Normal	Immune defence activation after 1st booster injection
Low	Low	Few significant effect
Normal	None	An immune defence is established in some cases
Normal	Normal	An immune defence is established in more cases
Normal	High	Targeted defence, but at the expense of other lymphocytes.
High	Low/Normal	Immune defence is established and persists well.
High	High	Targeted defence, but at the expense of other lymphocytes.

Wrap-up

Programming theory of biological processes can provide efficient

- Model building methods (modular, updates, testing, continuous integration, GitHub …)
- Model analysis methods (graph theoretic, abstract interpretation, model-checking, …)
- Before using classical mathematical analysis methods

Subcellular level: CRN programming language

BIOCHAM modeling platform

- Hierarchy of semantics ODE, CTMC, Petri Net, Boolean
- Explicit graph structure allowing for efficient analyses
  - Model comparison in the large by subgraph epimorphism
  - Graphical conditions for ODE conservation laws, rate-independence, basins of attraction
  - Graphical requirements for multistationarity [Baudier F- Soliman. Journal of Theoretical Biology, 459:79–89, 2018]
- CRN synthesis
  - Parameter optimization by adaptive evolution algorithm (AI intensification/exploration search)
  - CRN design by compilation (or artificial evolution) of function specification and comparison to natural CRNs

Multicellular, tissue, organ, body levels: multi-agent, algorithmic models

- Immune system: probabilistic parallel search reinforcement learning algorithm