Available online at www.sciencedirect.com
ScienceDirect
C. R. Biologies 329 (2006) 953–962
COMPTES RENDUS BIOLOGIES
http://france.elsevier.com/direct/CRASS3/

Biological modelling / Biomodélisation

Potential automata. Application to the genetic code III

Jacques Demongeot [a,b,*], Adrien Elena [a], Georges Weil [a]

[a] *TIMC-IMAG UMR CNRS 5525, faculté de médecine, université Joseph-Fourier, Grenoble, 38700 La Tronche, France*
[b] *Institut universitaire de France, 75005 Paris, France*

Received 3 March 2006; accepted after revision 10 July 2006

Available online 9 October 2006

Presented by Michel Thellier

Abstract

In previous notes, we have described both mathematical properties of potential (n-switches) and potential-Hamiltonian (Liénard systems) continuous differential systems, and also biological applications, especially those concerning primitive cyclic RNAs related to the genetic code. In the present note, we give a general definition of a potential automaton, and we show that a discrete Hopfield-like system already introduced by Goles et al. is a good candidate for such a potential automaton: it has a Lyapunov functional that decreases on its trajectories and whose time derivative is just its discrete velocity. Then we apply this new notion of potential automaton to the genetic code. We show in particular that the consideration of only physicochemical properties of amino-acids, like their molecular weight, hydrophobicity and ability to create hydrogen bonds suffices to build a potential decreasing on trajectories corresponding to the synonymy classes of the genetic code. Such an 'a minima' construction reinforces the classical stereochemical hypothesis about the origin of the genetic code and authorizes new views about the optimality of its synonymy classes. ***To cite this article: J. Demongeot et al., C. R. Biologies 329 (2006).***

Résumé

Automates cellulaires potentiels. Application au code génétique. Dans les notes précédant cet article, nous avons décrit des systèmes différentiels continus potentiels (de type « n-switches ») et potentiel-hamiltoniens (de type « systèmes de Liénard »), dont les équations étaient bien adaptées à la modélisation des systèmes dynamiques en biologie. Par exemple, un système différentiel potentiel défini sur $\Re^n$ est du type : $\forall i = 1, \ldots, n$, $\mathrm{d}x_i/\mathrm{d}t = -\partial P/\partial x_i$, où P est une fonction réelle continûment différentiable sur $\Re^n$. Nous donnons dans cet article une définition du même type pour les automates discrets et nous montrons, à titre d'exemple, un système de type Hopfield, déjà étudié par Goles et al., pour lequel il existe une fonction de Lyapunov décroissant le long des trajectoires. Nous montrons que la dérivée temporelle de cette fonction de Lyapunov est exactement la vitesse discrète de l'automate et nous appliquons cette nouvelle notion au code génétique. Nous montrons en particulier que l'on peut construire un potentiel uniquement à partir de propriétés physicochimiques simples des amino acides, comme leur poids moléculaire, leur hydrophobicité et leur capacité à créer des liaisons hydrogène, les orbites de l'automate potentiel correspondant n'étant autres que les classes de synonymie du code. Leur construction « a minima » renforce l'hypothèse classique stéréochimique quant à l'origine du code génétique et ouvre de nouvelles perspectives vers l'optimalité de ses classes de synonymie. ***Pour citer cet article : J. Demongeot et al., C. R. Biologies 329 (2006).***

* Corresponding author.
E-mail address: Jacques.Demongeot@imag.fr (J. Demongeot).

1631-0691/$ – see front matter
doi:10.1016/j.crvi.2006.07.010

Keywords: Potential automata; Lyapunov functional; Genetic code; Synonymy classes; Stereochemical hypothesis

Mots-clés : Automate potentiel ; Fonction de Lyapunov ; Code génétique ; Classes de synonymie ; Hypothèse stéréochimique

1. Introduction

A discrete dynamical system has the same definition as the continuous ones [1–9]. It involves a flow function f defined on $E \times T$, where T is a discrete time space (in general N) and E a discrete state space ($\{0,1\}^n$ in the Boolean case and more generally a countable subset of $\Re_+^n$), $f(e,t)$ representing, for each state e and time t, the state reached after time t by the trajectory starting in state e at time 0. We denote in general $f(x(0),t)$ by $x(t) = (x_i(t))_{i=1,\ldots,n}$, which permits to have a coherent notation for all the states of a trajectory. The set of such states is called the orbit of $x(0)$. Following [10], we can now define the discrete time derivative for the state vector $(x_i(t))_{i=1,\ldots,n}$ of an automaton by: $\Delta x_i/\Delta t = (x_i(t+\Delta t) - x_i(t))/\Delta t$, which reduces to $x_i(t+1) - x_i(t)$, if $\Delta t = 1$. By using the same formula, we can also define:

- the space derivative: $\Delta f(x)/\Delta i = (f(x_{i+\Delta i}) - f(x_i))/\Delta i = f(x_{i+1}) - f(x_i)$, if $\Delta i = 1$
- the partial state derivative: $\Delta f(x)/\Delta x_i = (f(x_i + \Delta x_i) - f(x_i))/\Delta x_i$.

A discrete automaton is defined by a transition function F: $\Delta x_i/\Delta t = (x_i(t+\Delta t) - x_i(t))/\Delta t = F_i(x(t))$, where F_i depends only on coordinates $(x_j(t))_{j\in V(i)}$, $V(i)$ being a neighbourhood of i in the space set S (in general the Manhattan – or L_1 – unit ball of $S \subseteq Z^m$ centred on i and having a radius equal to 1), with conditions defining V on the boundary of S (e.g., periodic) and with constraints on the discrete velocity ensuring that the flow remains in E.

The aim of the paper is to find a discrete analogous for the definition of a potential (or gradient) discrete dynamical system (called here potential automaton) similar to those known in continuous systems. Then we will study as biological application a potential automaton whose orbits are the synonymy classes of the genetic code, the potential deriving only from simple physicochemical properties of the amino-acids corresponding to these synonymy classes.

2. Definition of a potential automaton

A continuous potential differential equation on $\Re^n$ is defined by: $\forall i = 1,\ldots,n$, $\mathrm{d}x_i/\mathrm{d}t = -\partial P/\partial x_i$, where P is a real continuously differentiable function (e.g., a polynomial with real coefficients) on $\Re^n$. In the same way, a potential automaton on the discrete state space E is defined by:

$$x_i(t+1) = h\big(-\Delta P/\Delta x_i + x_i(t)\big) \tag{1}$$

where P is a real function (e.g., a polynomial with real coefficients) on E and h a function from $\Re$ to E, with boundary conditions ensuring that the flow remains in E. For example, in the Boolean case, we will choose for h the Heaviside function H: $H(s) = 1$ if $s > 0$, and $H(s) = 0$ if $s \leqslant 0$. In the integer case (E subset of N^n), h will be the identity, if P has integer coefficients and if $\forall i = 1,\ldots,n$, $\Delta x_i \in \{-1,0,1\}$. Then Eq. (1) rewritten as $\Delta x_i/\Delta t = h(-\Delta P/\Delta x_i + x_i(t)) - x_i(t)$ can be considered as the discrete equivalent of $\mathrm{d}x_i/\mathrm{d}t = -\partial P/\partial x_i$.

Example. In the Boolean case, if $P(x) = \sum_k ({}^t x A_k x) x_k + {}^t x W x + Bx$, where $A = (a_{ijk})$ is an interaction tensor, with $A_k = (a_{ij})_k$ as marginal matrices and $a_{iii} = 0$, $W = (w_{ij})$ is an interaction matrix and $B = (b_i)$ a threshold line vector, we have, for the partial space derivatives of P:

$$\begin{aligned}\Delta P/\Delta x_i = {} & \sum_{j,k}(a_{ijk} + a_{jik} + a_{jki})x_j x_k \\ & + \sum_j (w_{ij} + w_{ji})x_j + b_i \\ & + \left[w_{ii} + \sum_{j\neq i}(a_{ijj} + a_{jij} + a_{jji})x_j\right]\Delta x_i\end{aligned} \tag{2}$$

Then the potential automaton associated to P is defined by:

$$\begin{aligned}\Delta x_i/\Delta t = \Delta x_i &= -\Delta P/\Delta x_i \\ &= -\sum_{j,k}(a_{ijk} + a_{jik} + a_{jki})x_j x_k \\ &\quad - \sum_j (w_{ij} + w_{ji})x_j - b_i \\ &\quad - \left[w_{ii} + \sum_{j\neq i}(a_{ijj} + a_{jij} + a_{jji})x_j\right]\Delta x_i\end{aligned}$$

Hence we have:

$$\Delta x_i = -\Bigg[\sum_{j,k}(a_{ijk}+a_{jik}+a_{jki})x_j(t)x_k(t) + \sum_j (w_{ij}+w_{ji})x_j(t)+b_i\Bigg] \Bigg/\Bigg[1+w_{ii}+\sum_{j\neq i}(a_{ijj}+a_{jij}+a_{jji})x_j(t)\Bigg]$$

and $x_i(t+1) = H(\Delta x_i + x_i(t)) = H(-\Delta P/\Delta x_i + x_i(t))$, where H is the Heaviside function. From (2), we derive:

$$x_i(t+1) = H\Bigg(-\Bigg[\sum_{j,k}(a_{ijk}+a_{jik}+a_{jki})x_j(t)x_k(t) + \sum_j (w_{ij}+w_{ji})x_j(t)+b_i\Bigg] \Bigg/\Bigg[1+w_{ii}+\sum_{j\neq i}(a_{ijj}+a_{jij}+a_{jji})x_j(t)\Bigg] + x_i(t)\Bigg) \tag{3}$$

3. Properties of a potential automaton

Proposition 1. *Let us suppose that the state space E equals N^n and that P is defined on E by: $\forall x \in E$, $P(x) = \sum_k ({}^t x A_k x)x_k + {}^t x W x + Bx$, where A, W, B are respectively an integer tensor, an integer matrix and an integer line vector. Let suppose also that: $\forall i = 1,\ldots,n$, $\Delta x_i \in \{-1,0,1\}$. Consider now the potential automaton defined by: $x_i(t+1) = -\Delta P/\Delta x_i + x_i(t)$, if $x_i(t) > 0$, and by boundary conditions $x_i(t+1) \geqslant 0$, if $x_i(t) = 0$, such that the flow remains in E.*

Then, if the tensor A is symmetrical with vanishing diagonal (i.e. if $\forall i, j, k = 1,\ldots,n$, $a_{ijk} = a_{ikj} = a_{kij} = a_{jki} = a_{jik} = a_{kji}$ and $a_{iik} = 0$), and if each sub-matrix on any subset J of indices in $\{1,\ldots,n\}$ of A_k and of W is non-positive with vanishing diagonal, P decreases on the trajectories of the potential automaton, for any mode of implementation of the dynamics (sequential, block sequential and parallel). Hence the stable fixed configurations of the automaton correspond to the minima of its potential P.

Proof. We have:

$$\begin{aligned} &P\big(x(t+1)\big) - P\big(x(t)\big) \\ &= 3\sum_k\sum_{i,j} a_{ijk}\Delta x_i \Delta x_j x_k(t) + \sum_{i,j,k} a_{ijk}\Delta x_i\Delta x_j\Delta x_k \\ &\quad + 3\sum_k\sum_{i,j} a_{ijk}x_i(t)x_j(t)\Delta x_k \\ &\quad + \sum_i\Bigg(\sum_{j\neq i}(w_{ij}+w_{ji})x_j(t)\Delta x_i \\ &\quad + \sum_{j\neq i} w_{ij}\Delta x_i\Delta x_j + 2w_{ii}x_i(t)\Delta x_i + w_{ii}\Delta x_i^2\Bigg) \\ &\quad + \sum_i b_i\Delta x_i \end{aligned}$$

In a potential automaton, we have $\forall i = 1,\ldots,n$, $\Delta x_i = -\Delta P/\Delta x_i$, hence:

$$\begin{aligned} \Delta x_i = &-\Bigg[\sum_{j,k}(a_{ijk}+a_{jik}+a_{jki})x_j x_k \\ &+ \sum_j (w_{ij}+w_{ji})x_j + b_i\Bigg] \\ &- \Bigg[w_{ii} + \sum_{j\neq i}(a_{ijj}+a_{jij}+a_{jji})x_j\Bigg]\Delta x_i \end{aligned}$$

(1) In the sequential updating, only one component x_k changes its value between t and $t+1$, hence:

$$\begin{aligned} &P\big(x(t+1)\big) - P\big(x(t)\big) \\ &= 3\Bigg(\sum_{i\neq k} a_{ikk}\Delta x_k^2 x_i(t) + \sum_{i,j} a_{ijk}x_i(t)x_j(t)\Delta x_k\Bigg) \\ &\quad + \sum_j (w_{kj}+w_{jk})x_j(t)\Delta x_k + w_{kk}\Delta x_k^2 + b_k\Delta x_k \\ &= \Delta x_k(-\Delta x_k) = -\Delta x_k^2 \leqslant 0 \end{aligned}$$

(2) In the block-sequential updating, let denote by J the subset of $\{1,\ldots,n\}$ reached by the sequential iteration. Then only the x_k's for k in J can change their value between t and $t+1$, hence:

$$\begin{aligned} &P\big(x(t+1)\big) - P\big(x(t)\big) \\ &= 3\sum_k \sum_{(i,j)\in J\times J} a_{ijk}\Delta x_i\Delta x_j x_k(t) \\ &\quad + \sum_{(i,j,k)\in J\times J\times J} a_{ijk}\Delta x_i\Delta x_j\Delta x_k \\ &\quad + 3\sum_{k\in J}\sum_{i,j} a_{ijk}x_i(t)x_j(t)\Delta x_k \\ &\quad + \sum_{i\in J}\sum_j (w_{ij}+w_{ji})x_j(t)\Delta x_i \\ &\quad + \sum_{(i,j)\in J\times J} w_{ij}\Delta x_i\Delta x_j + \sum_{i\in J} b_i\Delta x_i \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\in J} \Delta x_i \Bigg[3 \sum_{j,k} a_{ijk} x_j(t) x_k(t) \\
&\quad + 3 \sum_{j\in J,k} a_{ijk} \Delta x_j x_k(t) + \sum_{(j,k)\in J\times J} a_{ijk} \Delta x_j \Delta x_k \\
&\quad + \sum_j (w_{ij} + w_{ji}) x_j(t) + \sum_{j\in J} w_{ij} \Delta x_j + b_i \Bigg] \\
&= \sum_{i\in J} \Delta x_i \Bigg[-\Delta x_i \Bigg(1 + w_{ii} + 3 \sum_{j\neq i} a_{ijj} x_j(t) \Bigg) \\
&\quad + \sum_{j\in J} \Delta x_j \Bigg(w_{ij} + 3 \sum_k a_{ijk} x_k(t) \Bigg) \\
&\quad + \sum_{(j,k)\in J\times J} a_{ijk} \Delta x_j \Delta x_k \Bigg] \\
&= -\sum_{i\in J} \Delta x_i^2 (1 + w_{ii}) + \sum_{(i,j)\in J\times J} w_{ij} \Delta x_i \Delta x_j \\
&\quad + 3 \sum_{(i,j)\in J\times J,k} a_{ijk} \Delta x_i \Delta x_j x_k(t) \\
&\quad + \sum_{(i,j,k)\in J\times J\times J} a_{ijk} \Delta x_i \Delta x_j \Delta x_k \\
&= -\sum_{i\in J} \Delta x_i^2 (1 + w_{ii}) + \sum_{(i,j)\in J\times J} w_{ij} \Delta x_i \Delta x_j \\
&\quad + 3 \sum_{(i,j)\in J\times J,k\notin J} a_{ijk} \Delta x_i \Delta x_j x_k(t) \\
&\quad + 2 \sum_{(i,j,k)\in J\times J\times J} a_{ijk} \Delta x_i \Delta x_j x_k(t) \\
&\quad + \sum_{(i,j,k)\in J\times J\times J} a_{ijk} \Delta x_i \Delta x_j \big(x_k(t) + \Delta x_k \big)
\end{aligned}
$$

From the hypotheses of the present proposition and from the positivity of $x_k(t)$ and $x_k(t+1) = x_k(t) + \Delta x_k$, we deduce that $P(x(t+1)) - P(x(t)) \leqslant 0$. □

Proposition 2. *In the Boolean case, let us suppose that* $A = 0$, $P(x) = {}^t x W x + Bx$, *with* $w_{ii} = 1$ *and each sub-matrix on any subset* J *of indices in* $\{1, \ldots, n\}$ *of* W *is non-positive. Then* P *decreases on the trajectories of the potential automaton defined by* $x_i(t+1) = H(-\Delta P/\Delta x_i + x_i(t))$ *for any mode of implementation of the dynamics (sequential, block sequential, and parallel). This automaton is a Hopfield-like neural network, whose stable fixed configurations correspond to the minima of* P.

Proof. It is easy to check that from (2): $\Delta P/\Delta x_i = \sum_j (w_{ij} + w_{ji}) x_j + b_i + w_{ii} \Delta x_i$, and from (3) and from $w_{ii} = 1$:

$$
\begin{aligned}
x_i(t+1) &= H\big(-\Delta P/\Delta x_i + x_i(t)\big) \\
&= H\Bigg(-\Bigg[\sum_j (w_{ij} + w_{ji}) x_j(t) + b_i \Bigg] \\
&\qquad /[1 + w_{ii}] + x_i(t) \Bigg) \\
&= H\Bigg(-\sum_{j\neq i} (w_{ij} + w_{ji}) x_j(t) - b_i \Bigg)
\end{aligned}
$$

Then, for the block sequential iteration: $P(x(t+1)) - P(x(t)) = -\sum_{i\in J} \Delta x_i^2 (1 + w_{ii}) + \sum_{(i,j)\in J\times J} w_{ij} \times \Delta x_i \Delta x_j \leqslant 0$, the result coming from the non-positivity of the sub-matrices of W or from [9]. □

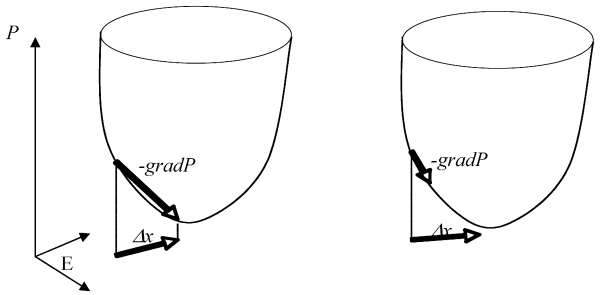

Fig. 1. Potential automaton with $\Delta x = -\operatorname{grad} P$ (on the left) and automaton with a Lyapunov function decreasing on its trajectories (on the right).

The interest of Proposition 2 is to show that the Hopfield-like automaton, defined by:

$$
\forall i = 1, \ldots, n,
$$

$$
x_i(t+1) = H\Bigg(-\sum_{j\neq i} (w_{ij} + w_{ji}) x_j(t) - b_i \Bigg)
$$

has not only P as Lyapunov function, as proved in [9], but more it can be considered as a potential automaton with a potential equal to P, because the opposite of the gradient of P is the velocity of the automaton, which is quite different in general for a system with simply a Lyapunov function (Fig. 1).

4. Application to the genetic code

We will now use the notion of potential automaton introduced above in order to show what kind of biological significance can have the attraction basins of the stable fixed configurations (i.e. corresponding to the minima of the potential) of such an automaton (see also [11]).

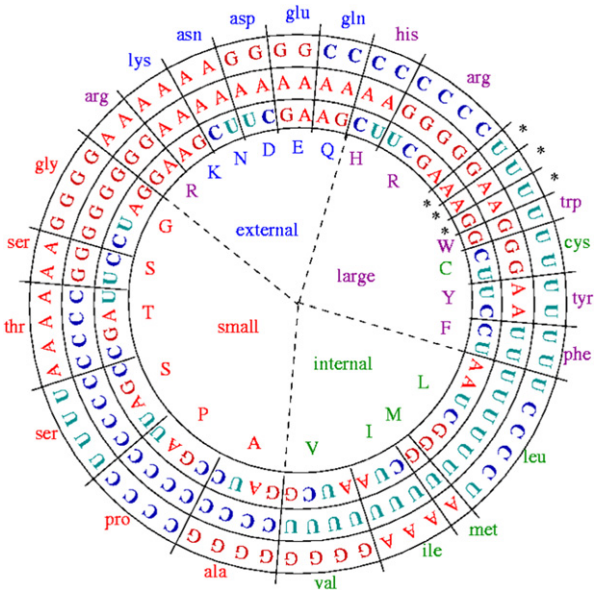

Fig. 2. Repartition of the amino-acids and of their codons following two criteria, size (small or large) and hydrophobicity (high/internal and low/external) (after [12]).

The stereochemical hypothesis belongs to a theory about the origin of the genetic code for which the first complexes between amino-acids (AA) and nucleic acids (NA) have been the start of the autopoietic origin of life. This theory claims that, from a physicochemical similarity on three (dependent) dimensions: (*i*) the ability L of binding (in particular through hydrogen bounds), (*ii*) the hydrophobicity H, and (*iii*) the molecular weight P, codons are associated in an unique and degenerate way to their AA [13,14]. This correspondence can be summarized in Fig. 2, in which the variables H, L and P are given for each amino-acid showing a structure into synonymy classes with sizes from 1 to 6.

According to the wobble hypothesis by Crick, the first two bases of a codon are essential for its affectation to a given amino-acid: we observe that the second base permits to share the codon space into classes (Fig. 3) associated with hydrophobic AA (central base uridine U or cytosine C) or heavy AA (central base U or adenine A), or able of binding AA (central base C or guanine G). Then we could represent each base by a Boolean number of three binary digits (e.g., 110 for U, the first 1 for the hydrophobicity, the second 1 pour the heavy weight and the last 0 for a weak ability of bind-

	U	C	A	G	
U	UUU Phenylalanine 2.8 3/7,165	UCU Serine *H*=–0.8 *L*=4/3,*P*=105	UAU Tyrosine – 1.3 4/7,181	UGU Cysteine 2.5 3/3,121	U
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine	C
	UUA Leucine 3.8 3/5,131	UCA Serine L=4/3	UAA Stop	UGA Stop	A
	UUG Leucine *P*>116	UCG Serine *P*<120	UAG Stop *P*>130	UGG Tryptophane –0.9 4/8,204	G
C	CUU Leucine 3.8 3/5,131	CCU Proline –1.6 3/4,115	CAU Histidine –3.2 5/6,155	CGU Arginine –4.5 6/6,174	U
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine	C
	CUA Leucine	CCA Proline *H*>0 *H*<0	CAA Glutamine –3.5 5/6,147	CGA Arginine	A
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine	G
A	AUU Isoleucine 4.5 3/5,131	ACU Threonine –0.7 4/4,119	AAU Asparagine – 3.5 5/5,132	AGU Serine –0.8 *L*=4/3 105	U
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine	C
	AUA Isoleucine *L*≤3/4	ACA Threonine *L*≤3/4	AAA Lysine –3.9 *L*≤5/5 4/6,146	AGA Arginine –4.5 6/6,174	A
	AUG Methionine/Start 1.9 3/5,149	ACG Threonine	AAG Lysine	AGG Arginine	G
G	GUU Valine 4.2 3/4,117	GCU Alanine 1.8 3/3,89	GAU Aspartate –3.5 5/4,133	GGU Glycine –0.4 3/3,75	U
	GUC Valine	GCC Alanine	GAC Aspartate	GGC Glycine	C
	GUA Valine	GCA Alanine	GAA Glutamate –3.5 5/5,147	GGA Glycine	A
	GUG Valine	GCG Alanine	GAG Glutamate	GGG Glycine	G

Fig. 3. Table of the genetic code, indicating the hydrophobicity H, the maximal number L of possible hydrogen bounds divided by the length of the longest carbonate chain and molecular weight P, for each amino-acid of a synonymy class of codons.

ing), summarizing their characteristics on the H, L and P axes. For the sake of simplicity, we will in the following reduce the coding to two binary digits, retaining only hydrophobicity and steric volume, redundant with P and L. We will show that it will suffice to explain the essential of the degeneracy of the genetic code, by taking into account optimality properties of the genetic code [15] and using a classical coding for the nucleic bases [1], i.e. U = 11, C = 10, A = 01 et G = 00. We will also define a potential automaton SUSY (SUrrogate SYstem), linking codons and AA having similar physicochemical properties like hydrophobicity $H \geqslant 0$ (resp. hydrophily $H < 0$) et big (resp. small) steric volume, positively correlated to P and L.

Let us define, for each configuration W in $\{0,1\}^6$, the set of Boolean numbers of six binary digits, the subset $V(W) = \{X;\, S(X,W) \geqslant 10 \text{ and } D(X,W) \leqslant 3/32\}$:

– where D is the Hamming distance, weighted following the Crick's wobble, which defines the codons ordering O (given in Fig. 3 from 1 to 64, for the Ws varying between 111111 and 000000):

$$\forall X, Y \in \{0,1\}^6$$
$$D(X,Y) = |x_3 - y_3| + |x_4 - y_4|/2 + |x_1 - y_1|/4 + |x_2 - y_2|/8 + |x_5 - y_5|/16 + |x_6 - y_6|/32,$$

– S is a similarity function between codons X and Y: $S(X,Y) = \sum_{i=1,5} a_i(x_i, y_i)x_i y_i + b_i(x_i, y_i)(1 - x_i)(1 - y_i)$,

– coefficients a_i and b_i are fixed by electrostatic (hydrophobicity H) and steric (molecular weight P and ability of binding L) properties of the aminoacids [16]:

$$a_1(1,1) = 2, \quad a_2(1,1) = 2, \quad a_3(1,1) = 3,$$
$$a_4(1,1) = 2, \quad a_5(1,1) = 2,$$
$$b_1(0,0) = 2, \quad b_2(0,0) = 3, \quad b_3(0,0) = 2,$$
$$b_4(0,0) = 3, \quad b_5(0,0) = 3.$$

Then the potential automaton SUSY is defined on $\{0,1\}^6$ by:

$$O\big(W(t+1)\big) = O\big(W(t)\big) + H\big(O\big(W(t) - O\big(\mathrm{Min}_O\, V(W(t))\big)\big)\big),$$

where for the ordering O, $O(W)$ is the rank of the Boolean number W and $\mathrm{Min}_O\, V(W(t))$ denotes the minimum of the subset $V(W(t))$, H being the Heaviside function. The potential P is defined by: $P(W) = O(W) - O(\mathrm{Min}_O\, V(W))$. P decreases along trajectories and has for attraction basins the first part of the states chains linked below by an arrow, the last state of a chain being a stable fixed configuration:

$(111111) \Rightarrow (111110), \quad (111101) \Rightarrow (111100)$
$(101111, 101110, 101101) \Rightarrow (101100)$
$(011111) \Rightarrow (011110) \quad (011101) \Rightarrow (011100)$
$(001111, 001110, 001101) \Rightarrow (001100)$
$(111011, 111010, 111001) \Rightarrow (111000)$
$(101011, 101010, 101001) \Rightarrow (101000)$
$(011011, 011010, 011001) \Rightarrow (011000)$
$(001011, 001010, 001001) \Rightarrow (001000)$
$(110111) \Rightarrow (110110), \quad (110101) \Rightarrow (110100)$
$(100111) \Rightarrow (100110), \quad (100101) \Rightarrow (100100)$
$(010111) \Rightarrow (010110), \quad (010101) \Rightarrow (010100)$
$(000111) \Rightarrow (000110), \quad (000101) \Rightarrow (000100)$
$(110011) \Rightarrow (110010), \quad (110001) \Rightarrow (110000)$
$(100011, 100010, 100001) \Rightarrow (100000)$
$(010011) \Rightarrow (010010), \quad (010001) \Rightarrow (010000)$
$(000011, 000010), (000001) \Rightarrow (000000).$

The attraction basins can be identified in the framework of the genetic code table in Fig. 4.

If we except codon 111100, whose similarity $S = 10$ with 101100, the only codons having an incorrect affectation [17,18] are:

- 011101, to relate to the block $(011111, 011110)$ $(S = 9)$,
- 110001, to relate to the block $(110101, 110100)$ $(S = 9)$,
- 010000, whose block has to be related to those of 100000 $(S = 8)$,
- 010011, whose block has to be related to those of 111011 $(S = 7)$.

For the two last codons above, the three digits coding with a complement of similarity equal to 1 for the equality between the third digit of the two first bases of a codon would relate 010001001 to the block of 101001001 $(S = 9)$ and 010001110 to the block of 110101110 $(S = 9)$. In any case of incorrect affectation we would be then very near to the above critical threshold of 10 defined for the similarity between the five first digits, in coherence with the stereochemical hypothesis and with the previous observations of the Gray code regularities [12,19,20].

	U	C	A	G
U	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine
	UUA Leucine	UCA Serine	UAA Stop	UGA Stop
	UUG Leucine	UCG Serine	UAG Stop	UGG Tryptophane
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine
	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine
	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine
	AUG Methionine/Start	ACG Threonine	AAG Lysine	AGG Arginine
G	GUU Valine	GCU Alanine	GAU Aspartate	GGU Glycine
	GUC Valine	GCC Alanine	GAC Aspartate	GGC Glycine
	GUA Valine	GCA Alanine	GAA Glutamate	GGA Glycine
	GUG Valine	GCG Alanine	GAG Glutamate	GGG Glycine

Fig. 4. Identification of the attraction basins of the automaton SUSY in the genetic code table.

AA	Human genome frequency	Codon frequency	Consensus chronology	Miller's order
Leu	99.1 1–	6/64	8	8
Ser	79,6 1–	6/64	7	7
Ala	71,0 1–	4/64	2	2
Glu	70,1 1–	2/64	5	5
Gly	66,6 1–	4/64	1	1
Val	61,5 1–	4/64	3	3
Pro	61,1 1–	4/64	6	6
Lys	57,0 1–	2/64	12	-
Arg	56,8 1–	6/64	10	-
Thr	53,1 1–	4/64	9	9

Fig. 5. Table giving the frequencies of occurrences of the AA in human genome [23] and in the genetic code, and giving their rank in a consensus chronology [22] and in the Miller experiment [22].

5. Variational properties of the system

The automaton SUSY not only maximizes a similarity function, but it gives final synonymy classes verifying two other variational properties, one based on a maximal information principle and the second showing a maximal congruence with present nucleic 'fossils' existing in many types of cells.

5.1. A maximal information principle

The easy way shown above to obtain a sketch of the synonymy classes through the potential automaton SUSY reinforces the stereochemical hypothesis. Apart from the amino-acids with six codons, the result is particularly convincing for the amino-acids the most frequent both in the Miller experiment and in the human genome as recalled in the table in Fig. 5 [21,22]. The mutual benefit taken by the RNA and AA worlds from a possible primitive direct association as predicted by the stereochemical theory (four out of the AA in Fig. 5, Leu, Glu, Val and Arg have preferential affinities with their codons and anti-codons [24]) is compatible with their Darwinian co-evolution well summarized by de Duve [25]: "The theory considered most likely today supposes a historical, co-evolutionary process in which the anticodons and the corresponding amino-acids were progressively recruited together under the control of natural selection. Several arguments support this hypothesis. The most convincing lies in the structure of the code, which, far from being random, happens to be such

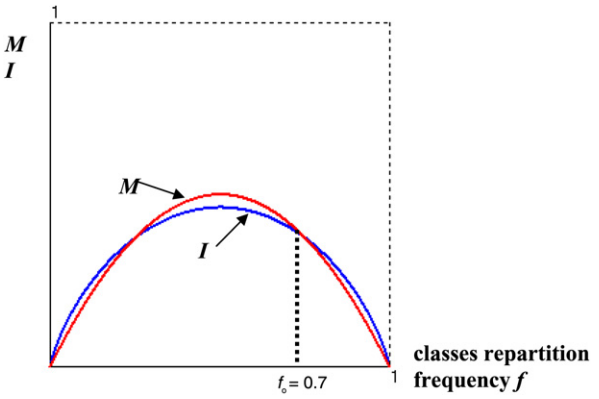

Fig. 6. Graph of the mutation function M and of the information I showing the optimal frequency f_o.

as to minimize the deleterious consequences of mutations."

We can interpret in a very schematic way this progressive adaptation by using a simple variational criterion, based on the dual principle of the minimization of the mutation function M and of the maximization of the genetic code information I.

Fig. 6 gives the optimal classes repartition frequency $f_o \approx 2/3$ for a code with 2 AA, obtained at the intersection of the graphs of the functions

$$M(f) = 2f(1-f) \quad \text{and}$$
$$I(f) = \big(-f\ln f - (1-f)\ln(1-f)\big) \times \int_0^1 M(x)\,dx \Big/ \int_0^1 H(x)\,dx,$$

where $H(x) = -x\ln x - (1-x)\ln(1-x)$. The frequency of an amino-acid AA_i class being equal to $n_i/64$, where n_i is its size, the mutation function $M = E/64$ is the expectation $E = \sum_{i=1,2}(64-n_i)n_i/64$ of the deleterious mutations (causing the exit of a synonymy class) divided by the number of codons. The information I is equal to the entropy of the distribution of the AA classes frequencies normalized such as M and I have the same mean value on [0, 1].

If now we renormalize the repartition of the genetic code in 21 synonymy classes or clusters: 3 of 6 codons, 5 of 4, 2 of 3, 9 of 2 and 2 of 1, following at each step the rule such as we roughly respect the optimal proportion $f_o \approx 2/3$ for one of the bigger renormalized class, we get a coherent tree (Fig. 7), satisfying at each bifurcation the above variational criterion. For example, the first renormalization concerns clusters of size 1 and 2, we combine for obtaining clusters of size 3 and 4 (see the bottom line of the tree) and secondly with clusters

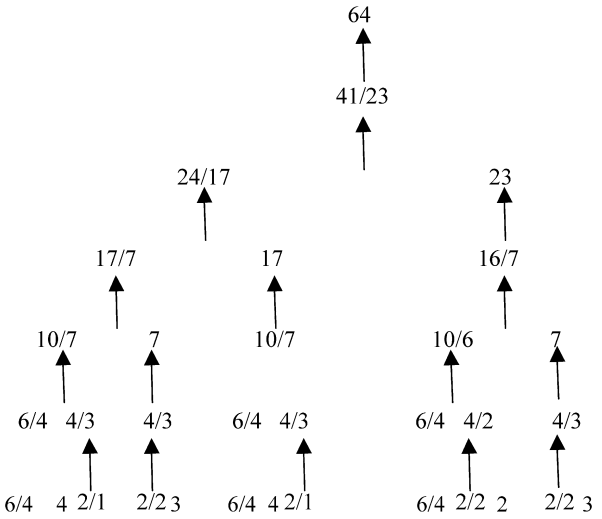

Fig. 7. Renormalization tree respecting the optimal frequency f_o for the synonymy classes of the genetic code.

of size 4 and 6 for obtaining the third line from the bottom, which corresponds to the affectation of codons to only 8 amino-acids (e.g., the 8 first in Miller's experiment). This short proof explains the classical properties of resistance to the mutations shown by the genetic code [26,27] compatible with the stereochemical hypothesis used in the automaton SUSY.

5.2. A maximal congruence with present nucleic 'fossils'

Some RNA relics exist in practically all eukaryote and prokaryote cells: tRNAs and miRNAs represent such fossils preserved during the evolution. We have already proved [1,2] that the most conserved part of tRNAS (their loops) were similar to primitive circular RNAs made of 1 and only 1 representative of each codon class, obtained by maximizing their affinity to AAs and minimizing their length. In the presently known functional miRNAS, we have the same similarities. For example, there is 222 known human miRNAs, plus 76 supplementary human miRNAs recently discovered [28], similar to a primitive circular RNA called AB: CAAGACUAUGAAUGGUGCCAUU, with a significativity p less than 21 – as shown in Fig. 8, which presents consensus sequences for these 76 miRNAs.

Fig. 8 presents also consensus sequences for miARNs coming from bacteria and plants [29,30] showing the same similarities. Further statistical studies [31] confirm this fact and the miRNA world can then be considered as a fossil reservoir coming from primitive RNAs built with respect to the synonymy classes obtained from the automaton SUSY.

U(UU)**UGC**(C)**A**(CC)**UU**(UUG)**GA**(G)**UGAA hsa miR 508 chromosome 19**
13/22 $p < 2 \times 10^{-3}$ (family of 7 miRNAs)
AAUGGU(U)**CC**(C)**UU**(U?)**AGA**(G)**U(**G)**U**(U) **hsa miR 522 chromosome X**
15/22 $p < 4 \times 10^{-4}$ (family of 37 miRNAs)
(U)**ACU**(CA)**G**(G)**A**(GA)**GUG**(G)**CA**(A)**UCA**(C) **hsa miR 510 chromosome X**
13/22 $p < 2 \times 10^{-3}$ (family of 3 miRNAs)
UUCAAG(C)**C**(?)**A**(?)**G**(G?)**GG**(?)**GC(**GU)**UU hsa miR 498 chromosome 19**
15/22 $p < 4 \times 10^{-4}$ (family of 29 miRNAs)

cau(gugu)**gacu**(ga)**g**(u)**a**(u)**uggug** miRNA SgrR **14/22** $p < 10^{-3}$
cc(c)**uu**(g)**aag**(u)**c**(ag)**ugaa** miRNA OxyS **12/17** $p < 2 \times 10^{-3}$
augaa(g)**gg**(a)**gc**(u)**c** miRNA AtmiR 159 **10/13** $p < 4 \times 10^{-3}$ (family of 4 miRNAs)
ugaa(g)**ggugc** miRNA 319 Saccharum oficinarum **9/10** $p < 6 \times 10^{-3}$
uu(g)**aagacua** miRNA AtmiR 161 **9/10** $p < 6 \times 10^{-3}$
aagacu(ug)**ga** miRNA PtmiR 163 **8/10** $p < 5 \times 10^{-2}$

Fig. 8. Similarities between miRNA consensus sequences and a circular primitive RNA called AB.

6. Conclusion

We have introduced in this paper the notion of potential automaton, which gives a natural framework to previous discrete systems like Hopfield neural networks, well studied in the past for the existence of Lyapunov energy functions. In a further study, we will introduce the notion of Hamiltonian automaton and propose new directions of research towards the potential-Hamiltonian decomposition of discrete dynamical systems (cf. [5,6] for the continuous case). An application of this new notion of potential automaton concerns the genetic code for which we proved that the degeneracy in synonymy classes of codons can be partly explained by the action of a potential automaton based on the stereochemical properties of both the amino-acids and their codons.

Acknowledgements

We are very indebted to L. Segel and J. Maynard-Smith (died in 2004) for inviting us to pursue researches in the field of theoretical biology and for showing and signalling the road.

References

[1] J. Demongeot, J. Besson, Code génétique et codes à enchaînement I, C. R. Acad. Sci. Paris, Ser. III 296 (1983) 807–810.
[2] J. Demongeot, J. Besson, Genetic code and cyclic codes II, C. R. Acad. Sci. Paris, Ser. III 319 (1996) 520–528.
[3] J. Demongeot, F. Leitner, Compact set valued flows I: Applications in medical imaging, C. R. Acad. Sci. Paris, Ser. IIb 323 (1996) 747–754.
[4] J. Demongeot, P. Kulesa, J.D. Murray, Compact set valued flows II: Applications in biological modelling, C. R. Acad. Sci. Paris, Ser. IIb 324 (1997) 107–115.
[5] J. Demongeot, N. Glade, L. Forest, Liénard systems and potential-Hamiltonian decomposition. I: Methodology, C. R. Mathématique, in press.
[6] N. Glade, L. Forest, J. Demongeot, Liénard systems and potential-Hamiltonian decomposition. II Applications in biology, C. R. Biologies, in press.
[7] E. Goles, F. Fogelman Soulié, D. Pellegrin, Decreasing energy functions as a tool for studying threshold networks, Discrete Appl. Math. 12 (1985) 261–277.
[8] F. Fogelman Soulié, E. Goles, S. Martinez, C. Mejia, Energy function in neural networks with continuous local functions, Complex Syst. 3 (1989) 269–293.
[9] M. Cosnard, E. Goles, Discrete states neural networks and energies, Neural Networks 10 (1997) 327–334.
[10] F. Robert, Discrete Iterations: A Metric Study, Springer, Berlin, 1986.
[11] O. Cinquin, J. Demongeot, High-dimensional switches and the modeling of cell differentiation, J. Theor. Biol. 233 (2005) 391–411.
[12] R. Swanson, A unifying concept for the amino acid code, Bull. Math. Biol. 46 (1984) 187–203.
[13] J.E. Blalock, K. Bost, Binding of peptides that are specified by complementary RNAs, Biochem. J. 234 (1986) 679–683.
[14] S. Mitaku, T. Hirokawa, T. Tsuji, Amphiphilicity index of polar amino-acids as an aid in the characterization of AA preference at membrane–water interfaces, Bioinformatics 18 (2002) 608–616.
[15] D. Gilis, S. Massar, N. Cerf, M. Rooman, Optimality of the genetic code with respect to protein stability and amino-acids frequencies, Genome Biol. 2 (2001) 1–12.
[16] G. Trinquier, Y.H. Sanejouand, Which effective property of amino acids is best preserved by the genetic code?, Protein Eng. 11 (1998) 153–169.
[17] M. Magini, J.E.M. Hornos, A dynamical system for the algebraic approach to the genetic code, Braz. J. Phys. 33 (2003) 825–830.
[18] J.E.M. Hornos, L. Braggion, M. Magini, M. Forger, Symmetry preservation in the evolution of the genetic code, Life 56 (2004) 125–130.
[19] D. Bosnacki, H. ten Eikelder, P. Hilbers, Genetic code as a Gray code revisited, in: Proc. Int. Conf. on Mathematics and Engineering Techniques in Medicine and Biological Sciences METMBS'03, Las Vegas, NV, USA, CSREA Press, 2003, pp. 447–456.

[20] M. He, S. Petoukhov, P. Ricci, Genetic code, Hamming distance and stochastic matrices, Bull. Math. Biol. 66 (2004) 1405–1421.
[21] A. Sciarrino, A mathematical model accounting for the organization in multiplets of the genetic code, BioSystems 69 (2003) 1–13.
[22] E. Trifonov, J. Sussman, The pitch of chromatin DNA is reflected in nucleotide sequence, Proc. Natl Acad. Sci. USA 77 (1980) 3816–3820.
[23] M.K. Hobish, N.S.M.D. Wickramasinghe, C. Ponnamperuma, Direct interaction between amino-acids and nucleotides as a possible physico-chemical basis for the origin of the genetic code, Adv. Space Res. 15 (1995) 365–375.
[24] I. Majerfeld, D. Puthenvedu, M. Yarus, RNA affinity for molecular L-histidine; genetic code origins, J. Mol. Evol. 61 (2005) 226–235.
[25] C. de Duve, Life Evolving, Oxford University Press, Oxford, UK, 2002.
[26] A. Figureau, M. Pouzet, Genetic code and optimal resistance to the effects of mutations, Orig. Life 14 (1984) 579–588.
[27] J.M. Labouygues, New mathematical model of genetic code with passive resistance to mutations or buccion and complementary dynamic protective mechanisms against noise at genome level, Agressologie 17 (1976) 329–335.
[28] I. Bentwich, A. Avniel, Y. Karov, R. Aharonov, S. Gilad, O. Barad, A. Barzilai, P. Einat, U. Einav, E. Meiri, E. Sharon, Y. Spector, Z. Bentwich, Identification of hundreds of conserved and non-conserved human μRNAs, Nat. Genet. 17 (2005) 766–770.
[29] S. Gottesman, Micros for microbes: noncoding regulatory RNA in bacteria, Trends Genet. 21 (2005) 399–404.
[30] B.H. Zhang, X.P. Pan, Q.L. Wang, G. Cobb, T. Anderson, Identification and characterization of new plant μRNAs using EST analysis, Cell Res. 15 (2005) 336–360.
[31] J. Demongeot, A. Moreira, Fossil tRNAs and miRNAs, and ancestral RNAs, PLOS Biol., submitted for publication.