Available online at www.sciencedirect.com
SCIENCE DIRECT®
BioSystems 77 (2004) 11–23
BioSystems
www.elsevier.com/locate/biosystems

An algorithm for the study of DNA sequence evolution based on the genetic code

G.Ch. Sirakoulis [a], I. Karafyllidis [a,*], R. Sandaltzopoulos [b], Ph. Tsalides [a], A. Thanailakis [a]

[a] *Department of Electrical and Computer Engineering, Democritus University of Thrace, 671 00 Xanthi, Greece*
[b] *Department of Molecular Biology and Genetics, Democritus University of Thrace, 681 00 Alexandroupolis, Greece*

Received 9 September 2003; received in revised form 11 December 2003; accepted 24 February 2004

Abstract

Recent studies of the quantum-mechanical processes in the DNA molecule have seriously challenged the principle that mutations occur randomly. The proton tunneling mechanism causes tautomeric transitions in base pairs resulting in mutations during DNA replication. The meticulous study of the quantum-mechanical phenomena in DNA may reveal that the process of mutagenesis is not completely random. We are still far away from a complete quantum-mechanical model of DNA sequence mutagenesis because of the complexity of the processes and the complex three-dimensional structure of the molecule. In this paper we have developed a quantum-mechanical description of DNA evolution and, following its outline, we have constructed a classical model for DNA evolution assuming that some aspects of the quantum-mechanical processes have influenced the determination of the genetic code. Conversely, our model assumes that the genetic code provides information about the quantum-mechanical mechanisms of mutagenesis, as the current code is the product of an evolutionary process that tries to minimize the spurious consequences of mutagenesis. Based on this model we develop an algorithm that can be used to study the accumulation of mutations in a DNA sequence. The algorithm has a user-friendly interface and the user can change key parameters in order to study relevant hypotheses.
© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: DNA; Modeling; Algorithms; Bioinformatics; Quantum mechanics; Nanotechnology

1. Introduction

Modeling DNA sequence evolution efforts beg the question whether mutations happen absolutely at random. The DNA of complex mammals comprises about 10^9 bases, whereas life on earth is about 10^{17} s old. The evolution of such a great complexity within this relatively short time period hints that evolution may not be completely random, but may be determined by some rules instead (Schwefel, 2002). In this vein, the elementary self-replicating module is supposed to be a short peptide comprising about 32 amino acids (McFadden, 2000). Since there exist 20 different amino acids, there are 20^{32} possible peptide sequence permutations 32 amino acids long. At a striking paradox, according to the random evolution model, assuming random synthesis of at least one molecule of every single possible peptide in the primordial chemical environment, out of which those able to replicate survived, the total mass of all possible peptides should have weighted about 10^{18} kg,

* Corresponding author. Tel.: +30-25410-79548; fax: +30-25410-26947.
E-mail address: ykar@ee.duth.gr (I. Karafyllidis).

0303-2647/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2004.02.006

a quantity vastly exceeding total carbon mass of tropical forests (McFadden, 2000).

Recently, quantum-mechanical models of DNA evolution proposed that evolution is directed by quantum-mechanical mechanisms (Baake et al., 1997, 1998; Bieberich, 2000; Kirby, 2002; McFadden and Al-Khalili, 1999; Ogryzko, 1997). The aforementioned models are strongly supported by recent data indicating that quantum proton tunneling causes tautomeric transitions in base pairs resulting in mutations during DNA replication (Golo et al., 2002; Hjort and Stafstrom, 2001; Kryachko, 2002). A complete quantum-mechanical description of DNA remains elusive, because of the complexity of the processes and the complex three-dimensional structure of the molecule (Altaisky, 2000; Balazs, 2003; Patel, 2001). Furthermore, several important theoretical problems have to be addressed, such as the transition from the quantum to the classical regime through the process of quantum measurement. Quantum measurement in biological systems is a very difficult and controversial issue (Rosen, 1996).

Although an acceptable quantum-mechanical model of DNA evolution is still distant, there is an increasing demand for the study of its evolution, because it may allow predictions of mutations. In this work we have developed a quantum-mechanical description of DNA evolution and, following its outline, we have constructed a classical model for DNA evolution in which some aspects of the quantum-mechanical processes are supposed to be reflected on the genetic code. An algorithm is developed based on this model. The algorithm has a user-friendly interface and the user can change several of its parameters, in order to study various hypotheses concerning DNA evolution models.

2. Quantum-mechanical description and mathematical modeling of DNA evolution

DNA can be modeled as a one-dimensional lattice in each site of which one of the four bases: A, C, T and G may be bound. We define as an *evolution event* a base change in a site (mutation). In the case of non-sexual reproducing, single cellular organisms, if a DNA sequence is passed unaltered from one generation to the next, then no changes occur and the DNA does not evolve. DNA evolves if a change occurs in one or more of its lattice sites (bases), either during DNA replication or during the life of the individual carrying the DNA.

The *time step* in DNA evolution is the time interval between two *evolution events* hence, time flow is not uniform. Consider for example a non-sexual reproducing species with a lifespan of 1 year. Suppose that a mutation occurs now, the next one occurs in 9 years, the next one in 4 years, and the next one in 8 months. Then, the first time step represents 9 years of real time, the second 4 years and the third 8 months. As a consequence of this model, the DNA strand and the individuals passing it from one generation to the other may exist in different time scales thus, DNA evolution is time-wise distinct from the life of the individuals that carry it.

2.1. Quantum-mechanical description of DNA evolution

We have developed a quantum-mechanical description of DNA, which to the best of our knowledge appears for the first time in the literature. We believe that any quantum-mechanical description of DNA should be developed in accordance with the quantum computation model, enabling thus its simulation and study using the forthcoming quantum computers. Physical systems are characterized by an ensemble of interacting constituents and can be generally represented and studied by different algebras of operators. In quantum computing the information unit is the quantum bit (qbit) and several qbits form a quantum register. The quantum register state evolves in time as a result of the action of quantum-mechanical operators, which are known as quantum gates (Williams and Clearwater, 1998). These operators form an algebra. For example all one-qbit operators can be described by an algebra generated by the Pauli spin 1/2 operators. To simulate a physical system using quantum computers a connection between the system and the computer should be found. This connection is a transformation (usually an isomorphism) of the operator algebra describing the physical system to the operator algebra used in quantum computing (Somaroo et al., 1999; Yepez, 2002). To simulate a physical system using quantum computers one must find an operator algebra describing the physical system and the transformation connecting this algebra with the one used in quantum computing.

DNA is a physical system and an operator algebra representing its function and evolution should exist. Such an algebra has not been found yet, but it will probably develop as a result of studying the physical and chemical mechanisms of DNA function and evolution. In this framework, we model DNA as a quantum system comprising a number of four base-state quantum subsystems located at the sites of its one-dimensional lattice. Each quantum subsystem may be found in any of the four base-states:$|A\rangle$, $|C\rangle$, $|G\rangle$ and $|T\rangle$, or in any linear combination of them, according to quantum superposition. To perform calculations, the four base-states must be represented by numbers. Since there are only four base-states the most appropriate way of representing them by numbers is to correspond each one of them to a respective number of the quaternary number system:

$$A \to 0,\ C \to 1,\ G \to 2,\ T \to 3 \tag{1}$$

Since in the double DNA strand A binds with T and C with G, by choosing the above representation the sum of bases at each base pair in the double strand is 3.

After that, the state of the quantum subsystem at the ith lattice site at time step t, $|b_i^t\rangle$, is given by:

$$|b_i^t\rangle = c_{0,i}^t\,|0\rangle + c_{1,i}^t\,|1\rangle + c_{2,i}^t\,|2\rangle + c_{3,i}^t\,|3\rangle \tag{2}$$

where $c_{0,i}^t$, $c_{1,i}^t$, $c_{2,i}^t$ and $c_{3,i}^t$ are the probability amplitudes of the subsystem state to be $|0\rangle$, $|1\rangle$, $|2\rangle$ or $|3\rangle$ at time t, respectively. The probability amplitudes are generally complex numbers and the corresponding probability is given by their square. All probabilities must add up to 1:

$$|c_{0,i}^t|^2 + |c_{1,i}^t|^2 + |c_{2,i}^t|^2 + |c_{3,i}^t|^2 = 1 \tag{3}$$

For example, in the case where the base C is sure to be found at the ith site:

$$|c_{0,i}^t|^2 = 0,\ |c_{1,i}^t|^2 = 1,\ |c_{2,i}^t|^2 = 0,\ |c_{3,i}^t|^2 = 0 \tag{4}$$

The state of the quantum subsystem at the ith lattice site at time step t, $|b_i^t\rangle$, is a vector in a four-dimensional Hilbert space, and the quantum state of the DNA molecule at time step t, $|\psi(t)\rangle$, is the tensor product of all m subsystem states:

$$|\psi(t)\rangle = |b_0^t\rangle \otimes |b_1^t\rangle \otimes |b_2^t\rangle \otimes \cdots \otimes |b_i^t\rangle \otimes \cdots \otimes |b_m^t\rangle \tag{5}$$

As any quantum system, the DNA molecule evolves in time according to the Schrödinger equation:

$$ih\frac{\partial\,|\psi(t)\rangle}{\partial t} = \hat{H}\,|\psi(t)\rangle \tag{6}$$

where the operator $\hat{H}$ is the Hamiltonian of the quantum system. If the initial conditions are known, i.e. if the base sequence in the DNA molecule is completely known at some time, this time is time step 0 and the initial state is $|\psi(0)\rangle$. The state of the DNA molecule at some later time t is given by the solution of the Schrödinger equation:

$$|\psi(t)\rangle = \exp\left(-\frac{i}{h}\hat{H}\,t\right)|\psi(0)\rangle \Rightarrow |\psi(t)\rangle = \hat{U}(t)\,|\psi(0)\rangle \tag{7}$$

In the matrix formulation of quantum mechanics the operator $\hat{U}$ is a unitary matrix. The evolution of the quantum state can be performed in discrete time steps (Williams and Clearwater, 1998). In this case the second Eq. (7) takes the form:

$$|\psi(t+1)\rangle = U^t\,|\psi(t)\rangle \tag{8}$$

where U^t is the operator applied at time step t. In Eq. (8), t in U^t is an index, not a power. Fig. 1 depicts this evolution. Fig. 1a is a schematic representation of Eq. (8). At time step t the quantum state of the DNA molecule, given by Eq. (5), is in the left column. The operator acts on this state and the result of the operation action is the quantum state of the DNA molecule at time step $t+1$, shown in the right column. The operator U^t represents in the physical and chemical processes that cause the evolution and is the tensor product of local operators as shown in Fig. 1b. These local operators may act on one or more quantum subsystems. If no operator acts on a subsystem, this no-action is represented by the unit operator I, which corresponds to the 2×2 unit matrix. For example, the operator U^t of Fig. 1a decomposed in local operators as shown in Fig. 1b is given by:

$$U^t = U_1^t \otimes U_2^t \otimes I \otimes U_3^t \otimes \cdots \otimes I \otimes U_n^t \tag{9}$$

The symbol $\otimes$ denotes the tensor product. The quantum-mechanical model of DNA described above can be mapped on a quantum computer with cellular architecture, where each quantum subsystem is represented by two qbits (Karafyllidis, 2003).

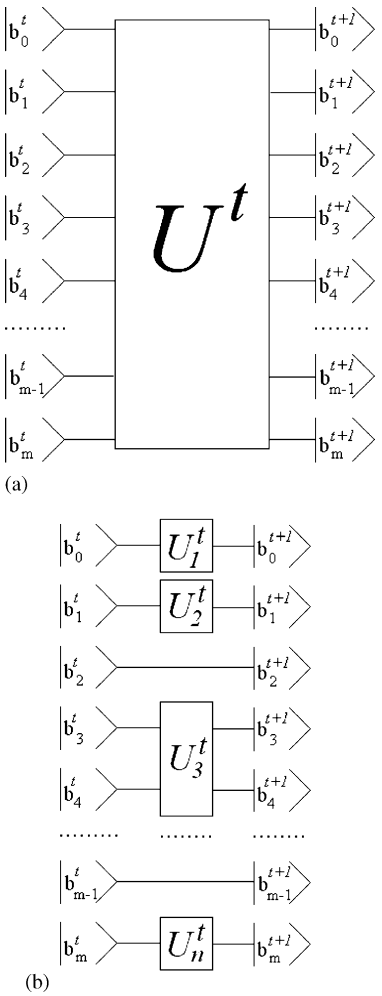

Fig. 1. Evolution of the quantum state of the DNA molecule in discrete time steps. (a) Evolution at time step *t*. (b) Decomposition of the operator U^t in local operators.

Two major problems in quantum-mechanical modeling of DNA evolution are now apparent. The first problem is the construction of the Hamiltonian and, consequently, of the unitary matrix representing U^t. The Hamiltonian must comprise all possible proton tunneling processes, the potential barriers that arise from the three-dimensional geometry of DNA and the charge distribution in it, which is not yet known. The second problem is the size of the unitary matrix representing U^t. The state vector of a DNA molecule with m bases exists in a 4^m-dimentional Hilbert space and the size of this matrix is $4^m \times 4^m$. It is obvious that the computational complexity for the simulation of DNA evolution is $O(4^{2m})$, where m is the number of bases in the molecule, and that it is an intractable *NP* (*non-polynomial*)-complete problem. This is not a new fact, since as a result of the hidden variable theorem no classical computer can simulate a quantum-mechanical system without suffering from exponential slowdown (Berman et al., 1998; Feynman, 1986).

Possibly, this problem will not be intractable for the forthcoming Quantum Computers (Berman et al., 1998; Feynman, 1986), but until they become available the quantum-mechanical problem described in this subsection must be reduced in order to be dealt with using classical computers.

2.2. Classical mathematical model of DNA evolution

Following the outline of the quantum-mechanical description of DNA sequence evolution, we developed a classical model in which the DNA is modeled as a one-dimensional lattice in each site of which one of the four bases A, C, T and G may be bound. The four bases are the states of the lattice sites and are represented by one of the numbers of the quaternary number system described in (1). The state of a DNA molecule with m bases is given by a vector which exists in an m-dimensional Cartesian space. For example, the state of the DNA molecule *ATCCGTT* that comprises seven bases is given by the one column vector $[0\ 3\ 1\ 1\ 2\ 3\ 3]^{\mathrm{T}}$. The state of a DNA molecule with m bases at time t, $[S^t]$, is given by:

$$[S^t] = \begin{bmatrix} b_1^t \\ b_2^t \\ b_3^t \\ \vdots \\ b_i^t \\ \vdots \\ b_m^t \end{bmatrix} \tag{10}$$

where b_i^t is the base located at the ith lattice site at time t and $b_i^t \in \{0, 1, 2, 3\}$. In analogy to Eq. (7) the

evolution of the DNA molecule with m bases from time step t to time step $t+1$ is given by:

$$[S^{t+1}] = \hat{M}^t [S^t] \tag{11}$$

where $[S^t]$ and $[S^{t+1}]$ are column vectors with m elements and $\hat{M}^t$ is an $m \times m$ matrix. The symbol t is an index denoting the time step. The full form of (11) is:

$$\begin{bmatrix} b_1^{t+1} \\ b_2^{t+1} \\ b_3^{t+1} \\ \vdots \\ b_i^{t+1} \\ \vdots \\ b_m^{t+1} \end{bmatrix} = \begin{bmatrix} M_{1,1}^t & M_{1,2}^t & M_{1,3}^t & \cdots & M_{1,i}^t & \cdots & M_{1,m}^t \\ M_{2,1}^t & M_{2,2}^t & M_{2,3}^t & \cdots & M_{2,i}^t & \cdots & M_{2,m}^t \\ M_{3,1}^t & M_{3,2}^t & M_{3,3}^t & \cdots & M_{3,i}^t & \cdots & M_{3,m}^t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M_{i,1}^t & M_{i,2}^t & M_{i,1}^t & \cdots & M_{i,i}^t & \cdots & M_{i,m}^t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ M_{m,1}^t & M_{m,2}^t & M_{m,3}^t & \cdots & M_{m,i}^t & \cdots & M_{m,m}^t \end{bmatrix} \begin{bmatrix} b_1^t \\ b_2^t \\ b_3^t \\ \vdots \\ b_i^t \\ \vdots \\ b_m^t \end{bmatrix} \tag{12}$$

The elements of the matrix $\hat{M}^t$ are real numbers and may be different in different evolution time steps. The element $M_{i,j}^t$ models the effect of the state at the jth site on the state at the ith site at the next evolution time step.

A comparison of the quantum-mechanical and the classical models shows that the complexity is very much reduced in the latter. The quantum-mechanical DNA state is given by a vector with 4^m elements existing in a 4^m-dimensional Hilbert space, whereas the classical DNA state is given by a vector with m elements which exists in a m-dimensional Cartesian space. The size of the unitary matrix representing $\hat{U}$ in (9) is $4^m \times 4^m$, whereas the size of the matrix $\hat{M}^t$ in (10) and (11) is $m \times m$. The complexity is significantly reduced, but all the information about DNA evolution that is stored in the quantum-mechanical state superposition is now inaccessible. The quantum-mechanical model is constructed within the Hilbert space, i.e. the elements of the state vector $|\psi(t)\rangle$ and the elements of the unitary matrix representing $\hat{U}$ are complex numbers. The classical model is constructed into the Cartesian space and the elements of the vector $[S^t]$ and of the matrix $\hat{M}^t$ are real numbers. Therefore, the information on the interaction between DNA bases, stored in the quantum-mechanical phase interference, is also inaccessible. Only a small part of the DNA evolution process can be described by any classical model, nonetheless classical models must be developed because of their importance for the study of DNA evolution. Furthermore, a good classical model may serve as a foundation for the development of more useful semi-classical models preserving a larger amount of information. The number of possible classical models is equal to the number of the matrices $\hat{M}^t$ that can be constructed. For example all models in which the elements of $\hat{M}^t$ are real numbers that are constant in time are linear models. A first-order approximation of a quantum model is a stochastic model. Following this line of thought we choose to model DNA evolution using a matrix $\hat{M}^t$ some elements of which change randomly in time according to a probability scheme given by the genetic code. Our model will be described in the following section.

3. A local stochastic classical model of DNA evolution

We start with the assumption that some of the quantum-mechanical processes in the DNA molecule are manifested in the structure of the genetic code (Osawa, 1995; Chechetkin, 2003). The genetic code is used to translate triplets of bases (codons) into amino acids, which are the protein building blocks. As a result of the redundancy in the genetic code, mutations at the third base of a codon are some times silent; when they occur, the new codon codes for the same amino acid. In this case the mutation has no or minimal biological consequences. On the other hand, mutations of the first base of a codon almost always result in a change in the amino acid encoded by the new codon, which frequently leads to disfunction and

even death. It is, therefore, reasonable to assume that the coding DNA sequences of the genome that has been selected by the evolution process are composed in such a way that renders mutations at the first or the second base of the codon less probable than mutations at the third base.

Our second assumption is that the occurrence of mutation at a DNA lattice site is affected by the charge distribution and the proton number near its neighborhood (Archilla et al., 2002). This assumption is modeled by setting equal to zero the elements of the $\hat{M}^t$ matrix that correspond to sites that are located in a distance greater than two lattice sites from the ith site. Thus, the state of the ith site (base) at the next evolution time step depends on the state of the two sites (bases) that are located nearest to it. The model is therefore local. Overlapping antiparallel Open Reading Frames (ORFs) are rare cases where our model cannot be applied as is. However, we foresee that suitable modifications of the model in future, may lead to a more complex version that could cover this exception, too. The values of the non-zero elements of the $\hat{M}^t$ matrix are not constant in time. Their values are determined according to the following evolution scheme:

1. The model starts with a given DNA sequence with m bases.
2. A real number, $0 \leq R < 1$, is chosen. This number is the mutation rate and the number of possible mutations, C, is given by:

$$C = m\,R \tag{13}$$

   If C is not an integer, then the real number C is rounded to the nearest smaller integer.
3. The number of bases that will likely mutate is C. At each evolution time step, a random number generator is used to generate C integers in the range from 1 to m. These numbers make a set S with cardinality C. The sites with numbers that belong to S are possible mutation sites. The rest of the sites will pass unchanged to the next DNA sequence:

$$\text{If } j \notin S \text{ then } [M^t_{i,j} = 0 \text{ for } i \neq j \text{ and } M^t_{j,j} = 1] \tag{14}$$

4. At each site that is candidate for mutation a G value is assigned, depending on the position of the site within the corresponding codon, according to Table 1 (a–c), respectively. These tables reflect

Table 1
We postulated that the mutation propensity at each nucleotide depends on the position of the site in the corresponding codon

	G	A	C	T	
(a)					
G	0	0	0	0	G
G	0	0	0	0	A
G	0	0	0	0	C
G	0	0	0	0	T
A	0.33	0	0	0	G
A	0.33	0	0	0	A
A	0	0	0	0	C
A	0	0	0	0	T
C	0.33	0	0	0.33	G
C	0.33	0	0	0.33	A
C	0	0	0	0	C
C	0	0	0	0	T
T	0	0	0	0.33	G
T	0	0	0	0.33	A
T	0	0	0	0	C
T	0	0	0	0	T
(b)					
G	0	0	0	0	G
G	0	0	0	0	A
G	0	0	0	0	C
G	0	0	0	0	T
A	0	0	0	0	G
A	0	0	0	0	A
A	0	0	0	0	C
A	0	0	0	0	T
C	0	0	0	0	G
C	0	0	0	0	A
C	0	0	0	0	C
C	0	0	0	0	T
T	0	0	0	0	G
T	0.33	0.33	0	0	A
T	0	0	0	0	C
T	0	0	0	0	T
(c)					
G	1	0.33	1	1	G
G	1	0.33	1	1	A
G	1	0.33	1	1	C
G	1	0.33	1	1	T
A	0.33	0.33	1	0.66	G
A	0.33	0.33	1	0	A
A	0.33	0.33	1	0.66	C
A	0.33	0.33	1	0.66	T
C	1	0.33	1	1	G
C	1	0.33	1	1	A
C	1	0.33	1	1	C
C	1	0.33	1	1	T
T	0	0.33	1	0.33	G
T	0	0.33	1	0.33	A
T	0.33	0.33	1	0.33	C
T	0.33	0.33	1	0.33	T

If a site is located at the first position of a codon, then its mutation probability is given by (a), at the second place by (b) and at the third place by (c). The mutation probability values found in the aforementioned tables result from the genetic code translating codons into amino acids.

Table 2
The universal genetic code (where *T* is used instead of *U*)

First position	Second position				Third position
	G	A	C	T	
G	Gly	Glu	Ala	Val	G
	Gly	Glu	Ala	Val	A
	Gly	Asp	Ala	Val	C
	Gly	Asp	Ala	Val	T
A	Arg	Lys	Thr	Met	G
	Arg	Lys	Thr	Ile	A
	Ser	Asn	Thr	Ile	C
	Ser	Asn	Thr	Ile	T
C	Arg	Gln	Pro	Leu	G
	Arg	Gln	Pro	Leu	A
	Arg	His	Pro	Leu	C
	Arg	His	Pro	Leu	T
T	Trp	STOP	Ser	Leu	G
	STOP	STOP	Ser	Leu	A
	Cys	Tyr	Ser	Phe	C
	Cys	Tyr	Ser	Phe	T

the mutation propensity for each site based on the genetic code shown in Table 2, where *T* is using instead of *U*. As a result, *G* values range, after normalization by factor 4 (i.e. the number of bases):

$$0 \leq G \leq 0.25 \tag{15}$$

In detail, if a site is located at the first or second position of a codon, then its mutation propensity is rather low, because the probability to generate a synonymous codon is low. Conversely, if a site is located at the third position of a codon, the propensity is rather high. For example, if the third position of the codon CTG (coding for leucine, Table 2) is mutated, there is 100% probability that a synonymous codon will be obtained, as shown in Table 1(c), and yielding a *G* value equal to 0.25. If the second position is mutated, it is absolutely sure that it will result in a different codon (Table 2), so it yields a *G* value equal to 0, as shown in Table 1(b). Finally, a mutation in the first nucleotide of this codon has a 33% chance to generate a synonymous codon, as shown in Table 1(c) and corresponds to a *G* value equal to 0.083. According to our assumption, this feature of the genetic code integrates, to a certain extend, the quantum-mechanical mechanism of DNA mutagenesis. Hence, we postulate that a mutation at the third position of the aforementioned codon is more probable than at the other two positions, exactly because the biological consequences of such an event are minimal.

5. The quantum-mechanical process that causes the mutation is modeled as a random process. A random number generator is used to assign another number *Q* at the site that is candidate for mutation,

$$0 \leq Q \leq 0.5 \tag{16}$$

It should be noted, that *Q* values are absolutely arbitrary. The higher the *Q* value's upper limit, the lower the impact of the codon position on the mutagenesis. Hence, manipulation of parameter *Q* is a built-in opportunity for fine-tuning the importance of nucleotide position within the codon on the process of mutagenesis. The two numbers *G* and *Q* are added and their sum is compared to a number, *P* that is a user-defined model parameter. If $P \leq (G + Q)$ mutation will occur, and if $P > (G + Q)$ will not.

6. Another set, *MUT*, is constructed. This set is a subset of *S* and contains all the sites at which mutation will occur. At the next evolution step, the states of the sites change to the values determined using a random number generator that for each mutation site *k* picks randomly one number, T_k, which is 0, 1, 2 or 3.

If $k \in \mathrm{MUT}$, then

$$\begin{cases} M^t_{i,k} = 0, & \text{if } i \neq k-2,\ k-1,\ k \\ M^t_{k-2,k} = 1, \quad M^t_{k-1,k} = 1,\ M^t_{k,k} = T_k \end{cases} \tag{17}$$

Thus, the next state at site *k* is given by:

$$b^{t+1}_k = M^t_{k-2,k}\, b^t_{k-2} + M^t_{k-1,k} b^t_{k-1} + T_k b^t_k \tag{18}$$

The addition in (18) is modulus 4. At each evolution time step the values of the elements of the matrix $\hat{M}^t$ are given by (14) and (17). Based on the model described above we developed an algorithm for DNA evolution which will be described in the next section.

4. DNA evolution algorithm

The flow chart of the algorithm is shown in Fig. 2. The algorithm starts by reading the mutation probabilities, i.e. Table 1. Then the user is prompt to enter

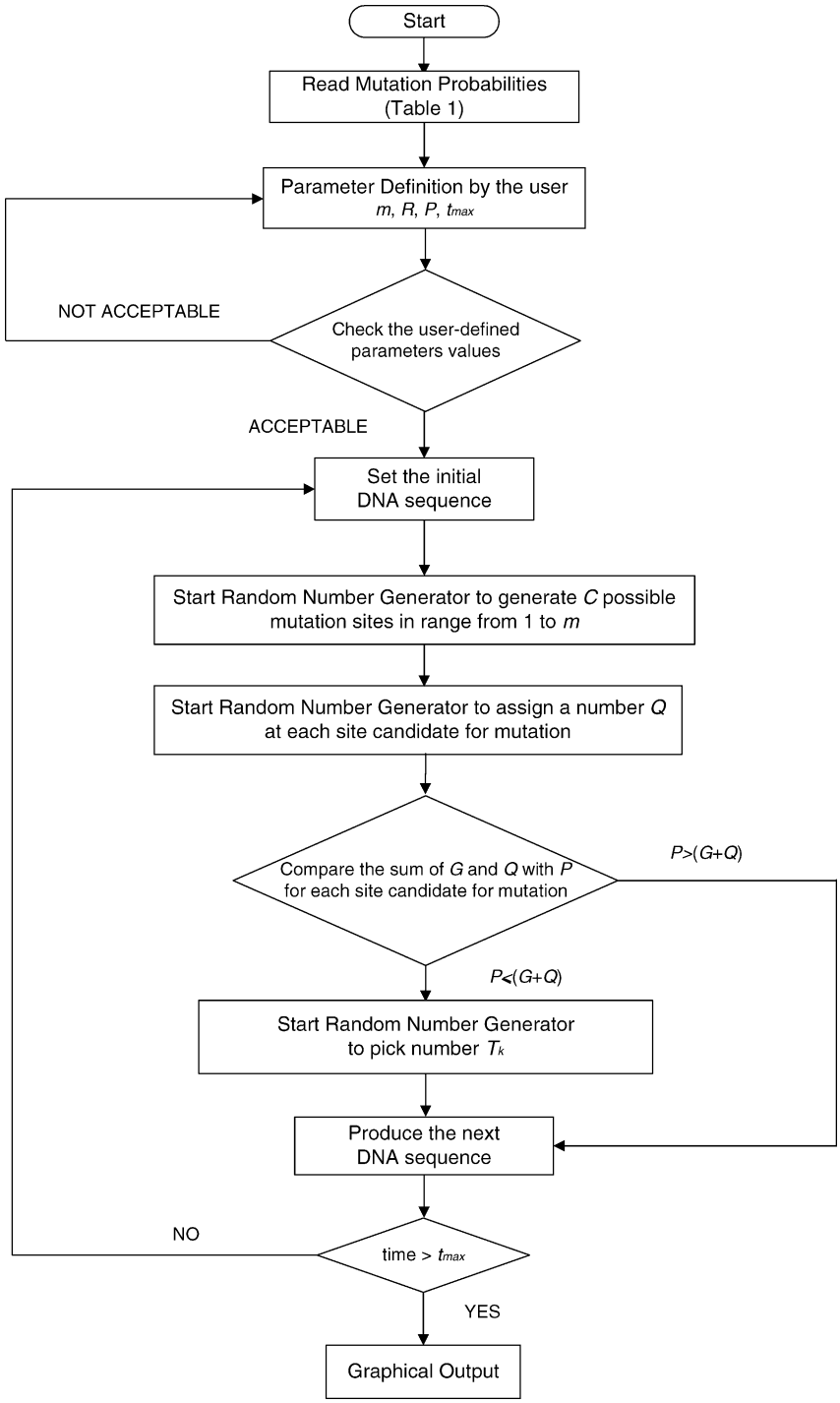

Fig. 2. The flowchart of the algorithm.

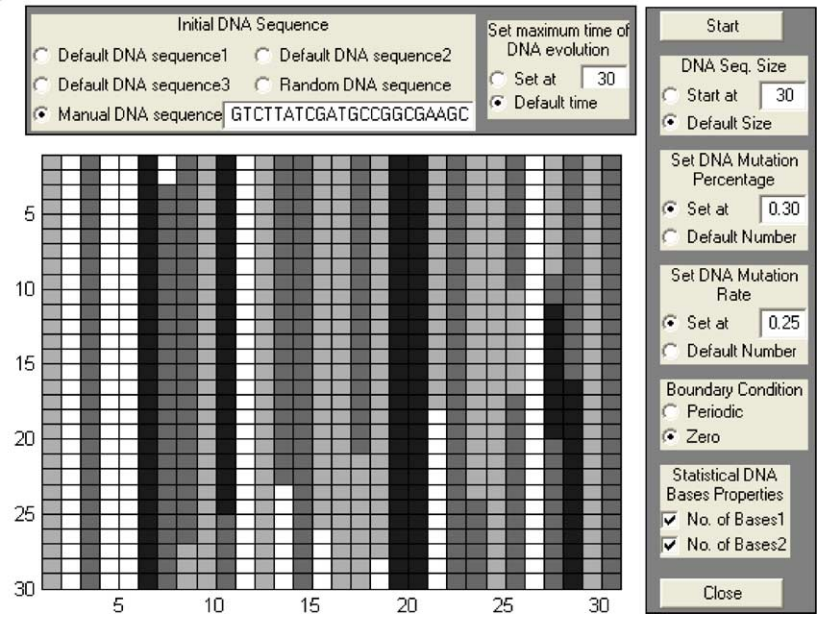

Fig. 3. The user graphical interface shows a DNA sequence evolution (A: black, C: dark gray, G: light gray, and T: white).

the values of the user-defined parameters, that is the length of the DNA sequence m, the mutation rate R, the number P which will be compared to the sum of G and Q, and t_{max}, the number of evolution time steps that the algorithm will take. The values of these parameters are checked and if not acceptable the user is prompted to correct them. Acceptable values for the parameters m and t_{max} must be positive integers and more specifically, m should be a multiple of 3, while R and P must be positive real numbers smaller than 1.

If the parameters are acceptable, the user determines the initial DNA sequence. The user may enter whichever sequence she/he wishes. Then the random number generator indicates the possible mutation sites and a random number is assigned to each site. Subsequently, P is compared to the sum $G + Q$ and, if it is greater, the site passes unaltered to the new sequence, otherwise the site state (base) is changed according to (18). The new DNA sequence is produced and, if the number of time steps taken is less than or equal to t_{max}, the algorithm executes another loop, otherwise a graphical output is produced and the algorithm stops.

The graphical user interface of the algorithm is shown in Fig. 3. In the field "Initial DNA Sequence" the user inserts the sequence which will be used as initial. The user has three choices: by clicking on the radio button beside the "Default DNA sequence 1, 2 or 3," a previously defined DNA sequence is used as initial. By clicking on the radio button beside the "Random DNA sequence," a randomly generated DNA sequence is used as initial. The user can enter hers/his own DNA sequence by clicking on the radio button beside the "Manual DNA sequence" and entering the sequence into the blank field on the right side of "Manual DNA sequence." The sequence entered by the user must be in frame with the corresponding ORF. That is to say, the first nucleotide of the input sequence must be the first nucleotide of a codon. In addition this sequence should not encompass any splice sites.

The number of evolution time steps, t_{max}, is entered in the field "Set maximum time of DNA evolution." The number m, i.e. the length of the DNA sequence is entered in the field "DNA Seq. Size." The numbers P and the mutation rate R, are entered in the fields "Set DNA Mutation Percentage" and "Set DNA Mutation Rate," respectively. The user can set periodic or zero

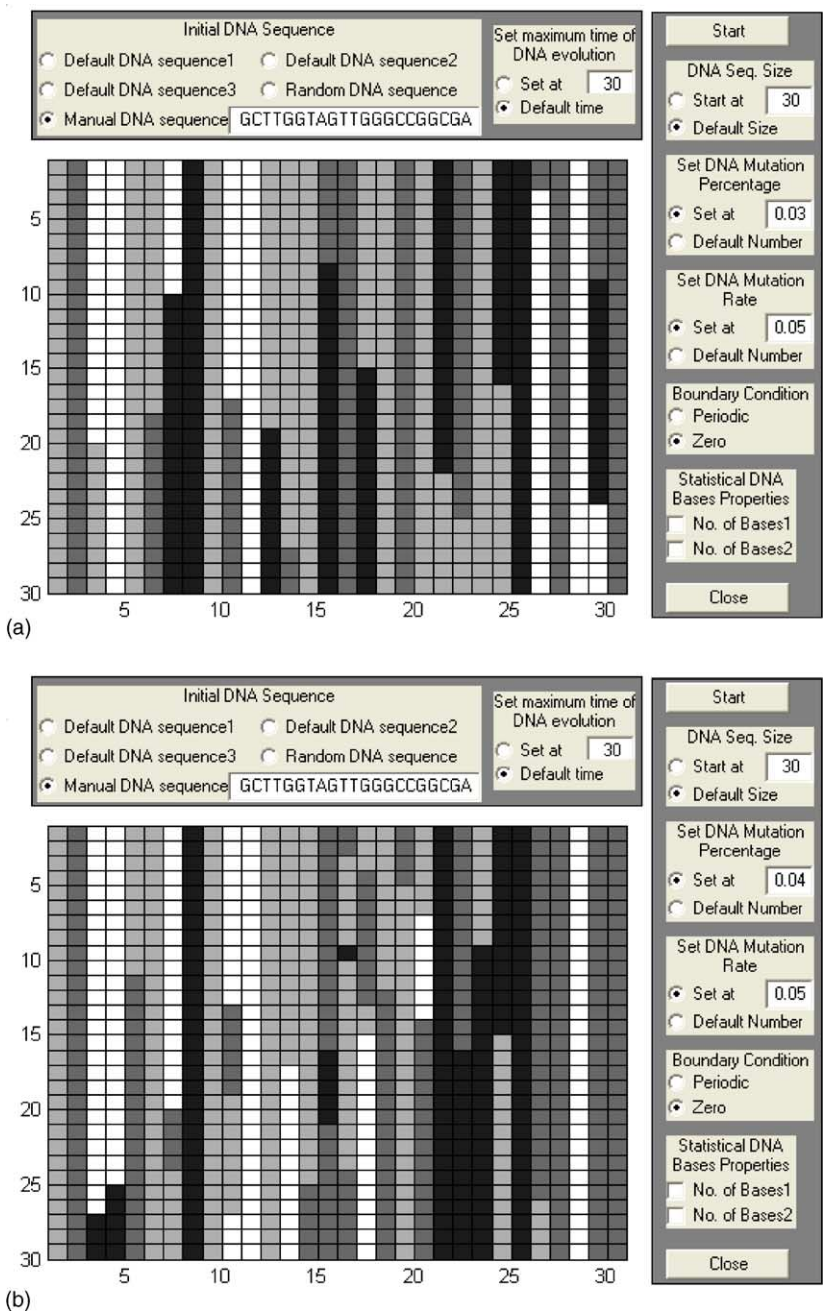

Fig. 4. (a) Application of the algorithm to a DNA sequence with $P = 0.03$, and (b) the same initial DNA sequence and the same parameters as in Fig. 3a are used except for the value of P which is set equal to 0.04 (A: black, C: dark gray, G: light gray, and T: white).

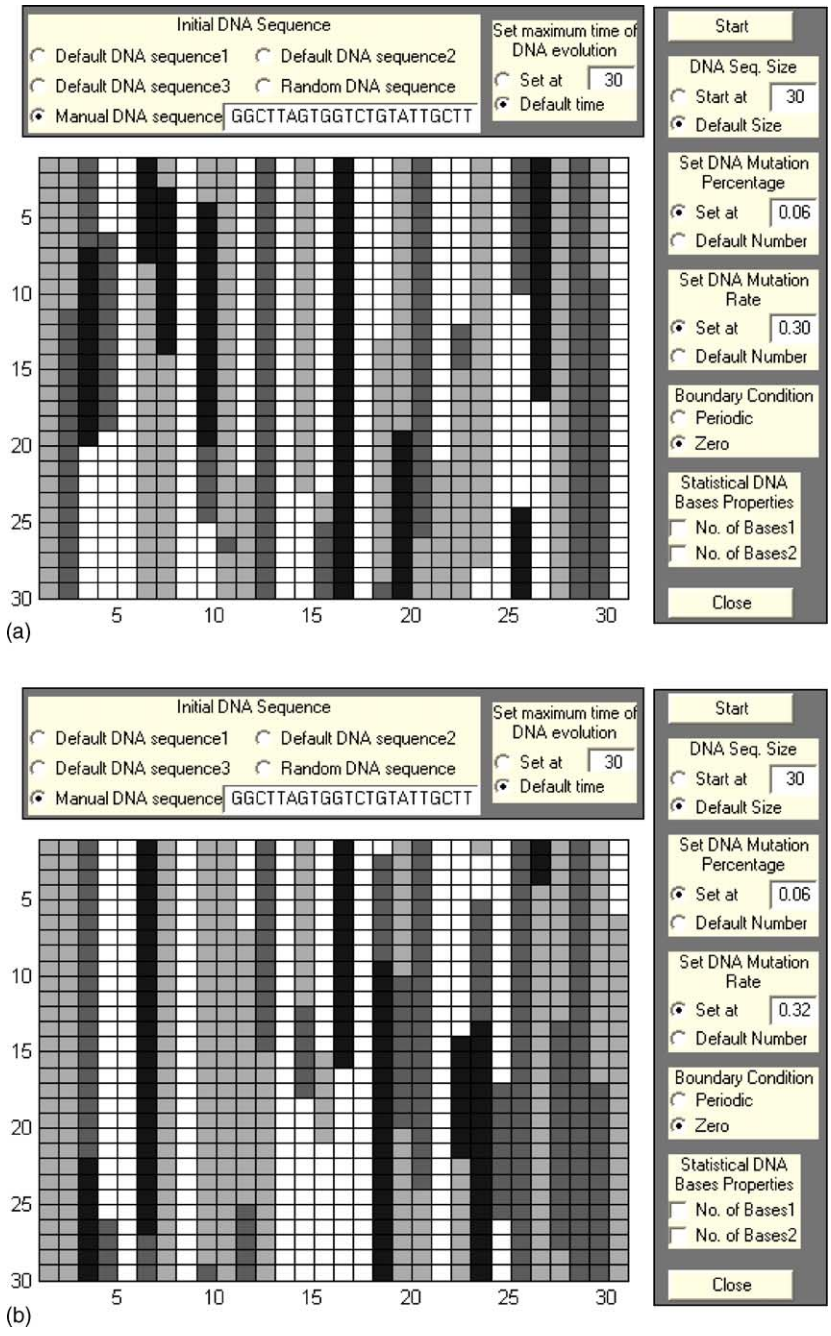

Fig. 5. (a) Application of the algorithm to a DNA sequence with $R = 0.30$, and (b) the same initial DNA sequence and the same parameters as in Fig. 4a are used except for the value of R which is set equal to 0.32 (A: black, C: dark gray, G: light gray, and T: white).

boundary conditions by clicking the corresponding radio button in the field "Boundary Condition." Some statistical properties, such as the number of bases at each evolution time step, as well as the difference between the number of each base in the beginning and at the end of the evolution, are displayed if she/he clicks on the check buttons "No. of Bases" in the "Statistical DNA Bases Properties" field. The graphical output is shown in the large box located down and left in the graphical user interface. The *x*-axis represents the sites of the DNA lattice and the *y*-axis the number of evolution time steps. The four bases are represented as follows: A: black, C: dark gray, G: light gray, and T: white.

The algorithm sensitivity on the two most important user defined parameters P and R is shown in Figs. 4 and 5. In Fig. 4a the algorithm is applied to a DNA sequence with $P = 0.03$. In Fig. 4b the same initial DNA sequence and the same parameters as in Fig. 4a are used except for the value of P which is set equal to 0.04. The non-linearity of the model is manifested by the significant change in the mutation pattern caused by a minute difference in the value of P. In Fig. 5a the algorithm is applied to a different DNA sequence with $R = 0.30$. In Fig. 5b the same initial DNA sequence and the same parameters as in Fig. 5a are used except for the value of R which is set equal to 0.32. Again the non-linearity of the model is obvious.

This algorithm with its graphical user interface is a useful computer tool for the study of various evolution hypotheses. The source code is free and is available upon demand.

5. Conclusions

We have developed a quantum-mechanical description of DNA evolution and, following its outline, we have constructed a classical local stochastic model for DNA evolution in which we assume that some aspects of the quantum-mechanical processes are manifested in the structure of the genetic code. Our second assumption is that the occurrence of mutation at some DNA lattice site is affected by the state of its neighbors, i.e. that site mutation is a local phenomenon. Based on this model we have developed an algorithm that can be used to study the DNA sequence evolution. The algorithm has a user-friendly interface and the user can change several parameters of the model, in order to study various evolution hypotheses.

References

Altaisky, M.V., 2000. What can biology bestow to quantum mechanics? Los Alamos preprint archive http://xxx.lanl.gov/abs/quant-ph/0007023.

Archilla, J.F.R., Christiansen, P.L., Gaididei, Yu.B., 2002. Interplay of nonlinearity and geometry in a DNA-related, Klein-Gordon model with long-range dipole–dipole interaction. Phys. Rev. E 65, 0166091–0166097.

Baake, E., Baake, M., Wagner, H., 1997. Ising quantum chain is equivalent to a model of biological evolution. Phys. Rev. Lett. 78, 559–562.

Baake, E., Baake, M., Wagner, H., 1998. Quantum mechanics versus classical probability in biological evolution. Phys. Rev. E 57, 1191–1192.

Balazs, A., 2003. On the physics of the symbol-matter problem in biological systems and the origin of life: affine Hilbert spaces model of the robustness of the internal quantum dynamics of biological systems. Biosystems 70, 43–54.

Berman, G.P., Doolen, G.D., Mainieri, R., Tsifrinivich, V.I., 1998. Introduction to Quantum Computers. World Scientific, London.

Bieberich, E., 2000. Probing quantum coherence in a biological system by means of DNA amplification. Biosystems 57, 109–124.

Chechetkin, V.R., 2003. Block structure and stability of the genetic code. J. Theor. Biol. 222, 177–188.

Feynman, R., 1986. Quantum mechanical computers. Found. Phys. 16, 507–531.

Golo, V.L., Volkov, Yu.S., 2002. Tautomeric transitions in DNA. Los Alamos preprint archive http://xxx.lanl.gov/abs/cond-mat/0110599.

Hjort, M., Stafstrom, S., 2001. Band resonant tunnelling in DNA molecules. Phys. Rev. Lett. 87, 228101–228110.

Karafyllidis, I., 2003. Cellular quantum computer architecture. Phys. Lett. A 320, 35–38.

Kirby, K.G., 2002. Biological adaptabilities and quantum entropies. Biosystems 64, 33–41.

Kryachko, E.S., 2002. The origin of spontaneous point mutations in DNA via Lowdin mechanism of proton tunnelling in DNA base pairs. Int. J. Quant. Chem. 90, 910–923.

McFadden, J., 2000. Quantum Evolution. Flamingo, Harper Collins Publishers, London.

McFadden, J., Al-Khalili, J., 1999. A quantum mechanical model of adaptive mutation. Biosystems 50, 203–211.

Ogryzko, V.V., 1997. A quantum-theoretical approach to the phenomenon of directed mutations in bacteria. Biosystems 43, 83–95.

Osawa, S., 1995. Evolution of the Genetic Code. Oxford University Press, Oxford.

Patel, A., 2001. Why genetic information processing could be quantum. J. Biosci. 26, 145–151.

Rosen, R., 1996. Biology and the measurement problem. Comput. Chem. 20, 95–100.

Schwefel, H.P., 2002. Deep insight from simple models of evolution. Biosystems 64, 189–198.
Somaroo, S., Tseng, C.H., Havel, T.F., Laflamme, R., Cory, D.G., 1999. Quantum simulations on a quantum computer. Phys. Rev. Lett. 82, 5381–5384.
Williams, C.P., Clearwater, S.H., 1998. Explorations in Quantum Computing. Springer-Verlag, Berlin.
Yepez, J., 2002. Quantum computation for physical modelling. Comput. Phys. Commun. 146, 277–279.