Simulation of DNA Strings Dynamics

Simonetta Astolfi and Patrizia Mentrasti
Dipartimento di Matematica
Università degli Studi di Roma "La Sapienza"
P.le Aldo Moro, 5, 00185 Roma, Italy
Email: astolfi@mat.uniroma1.it, mentrasti@mat.uniroma1.it

Abstract Many efforts are currently dedicated to DNA strings study. Here, we want to introduce a little improvement along the way of representing the microscopic behaviour of DNA molecules. Shortly, we obtained a computer model describing the movement of long or shortened DNA chains, when embedded in a liquid; indeed, we have strings made of monomers, which can be of different kinds and consequently move in locally chosen directions.
These monomers are forced by the rule to respect the "connectivity" and "excluded volume" constraints and to take into account behaviours coming from real life, since real sequences are given as initial configurations. We bring only a little part of the results deriving from a lot of computer hours of experiments and also give some ideas for further improvements.

Keywords: cellular automata, DNA strings, dynamics simulation.

1 Introduction

In recent years computers became capable of performing tasks that a short time ago could only by accomplished by men without achieving the same speed, efficiency or degree of precision. All these and other features made possible entangled complex systems dynamics simulations; so, many other new requests turned satisfiable. In our case, we can reproduce a microscopical phenomenon, that is DNA chains movement in a liquid, at equilibrium temperature, avoiding undesired denaturation effects; in fact, we provide a magnification of moving DNA strings, by studying the structure of molecules composing the deoxyribonucleic acid chains. Really, we use dedicated computers, so that speed of movement is not so far from the real one and the simulation results are dynamic, as it operates in parallel and one half of monomers move at the same time.

In these systems we do not want to provide a way of *studying* reality, because we do not analyze any physically existing matter; rather, we want to *represent* the real phenomenon, by reproducing it with theoretical studies and computing tools.

In a computer representation, we do not think that everything is a mechanical application; we want to put in evidence the human aspect in the automatical design. From the hardware point of view, a program is schematically read, compiled and executed; it is always the same procedure, but it gives us great advantages such as speed and the ability to repeat experiments. Behind all that, there is a concentrated human effort. We have to build a model from an idea and adapt it to the simulation. Then we design an algorithm and implement it on the machine. All this must be such that the output result is easily understood and the model respects all the bondary constraints.

Hence, our model has great intrinsic complexity, because DNA, into its structure, takes into account several elements interacting with each other and it is necessary to consider the multiplicity of their properties. We represent the primary structure of DNA, that is the sequencing of the four bases, adenine (A), thymine (T), guanine (G) and cytosine (C), and we also simulate the movement of only one chain of the double helix, moreover in two dimensions. We represent the double helix displacing the upper helix and hiding the lower one. The studied dynamics concern two kinds of nucleotides, as we do not distinguish between adenine and thymine and between guanine and cytosine too. Dynamics roles are obtained from intrinsic properties of these molecules, in particular from their atomic weight and the hydrogen bonds [**?**], which bind bases of the double helix.

Our algorithm takes inspiration from the model for monomeric chains by Bar-Yam [**?**] and it is an evolution of an implementation, made by Norman Margolus [**?**] on a dedicated machine [**?**], suitably designed for cellular automata simulations, that is for displacing rules for a particular abstract universal machine, named *cellular automaton*. We use the same machine [**?**] [**?**], by defining rules which distinguish among different bases and which choose the movement direction according to atomic weight of the bases. Besides, our algorithm, while simulating polymeric chains with two kinds of monomers, uses as sequence of bases real ones, taken from the ADH II (Alcohol dehydrogenase II) of Saccharomyces cervesiae, complete gene with regulating zone.

2 Behaviour of the model

In our simulations, one interesting feature is the speed of evolution, as the updating is performed at a rate of about nine steps per second; our model is very dynamic: the chains rapidly become different from the initial configuration and evolve very quickly.

In spite of all the necessary simplifications, mainly due to the two dimensional representation, our model acts in a way that there is a good accordance between reality and simulation. First of all, some unpredictable behaviour is reproduced by random generation; this means that examination of the algorithm implementation does not tell us which movement will be chosen by a particular monomer at every step. Actually, the exchange of two elements of the matrix used for random generation [**?**] or some other little modification, will make the monomers move differently.

Our model has an interesting macroscopic likelyhood: in fact images obtained from our simulations are very similar to those obtained from a real DNA (see section **??**).

In order to observe different behaviour related to diplacement and length of the chains, we constructed several different initial configurations. The first evident remark is a clear difference in behaviour between open and closed strings. The first ones, when placed *in a straight line*, lose their stucture as much as their initial position, dispersing themselves in the space; instead, the closed configurations, for example the *rhomb*, move always around their center of gravity, that never goes far from its initial position. We can think that this different behaviour depends on the additional degree of freedom of monomers with free extremities, as already conjectured in our previous simulations, which were obtained by a former model for monomeric chains [**?**]. In fact, the extremities seem to move first and more freely in the simulation, and it seems that they take away the whole chain. Because in the closed curves no component has this additional degree of freedom, their configuration remains stable around the center of gravity, as expected.

Another interesting fact is that, when comparing placed chains *in a straight line* with different lenght, the shorter ones lose their initial stucture faster then longer ones; probably, influence of the extremities acts later in the longer strings than in the shorter ones and movements of middle parts are narrowed by more limitations.

However, no straight string remains that way and, when observing many straight strings, some will writhe more than others, in any case it will be impossible to recognize initial configurations. In fact, when dis-

placing many straight strings in our space (essentially represented on the computer screen), some will tend to form entangled zones for a short time, others remain isolated and some empty side and some filled side will come. The alternation of these situations is very dynamic. It is interesting to compare these agglomerates. At any given movement chains look like they are joined in some points; after few steps the monomers of different strings, being before very near, go away. All this, stated by the entangled computer rule completely described in [**?**], is coherent with the real behaviour, where all this happens because repulsion forces act on monomers of different chains.

3 Interesting evolutions

Here we will not attempt to describe in the details the techniques of simulation of complex systems, particularly not the ones devoted to use of the cellular automata paradigm. There are many books on this subject and we refer to some of them [**?**], [**?**], and [**?**] for a wide description. In a similar way, we cannot even give some initial definitions of the primary structure of the DNA chain, together with the characteristics of the molecules and also for this item we refer to some books and papers like [**?**], [**?**], [**?**] and [**?**].

We only present some interesting features of a small part of the simulations we carried out. All of the experiments were carried out on the CAM-8 machine [**?**] owned by Department of Mathematics, University of Rome "La Sapienza".

CAM-8 is a cellular automata dedicated device, which has a Sparc Station as host and run by means of its own designed software; all this is difficult to understand and even to use, but suited at the purpose and very powerful.

An exhaustive description of the system which we implemented, together with the experiments and their complete description is in [**?**]. We only summarize the method we used to obtain the figures that appear in this paper.

3.1 Open strings

Now, we are going to show some configurations having DNA chain segments placed as open curves. The figures are here presented in this way: the first one is the initial configuration, having monomers taken

from real sequences; the second one is the configuration which appears after a defined number of steps. This number of steps is 250.000 in the first experiment that we present here and 100.000 in the others.

Different monomers are represented with different colors, in particular:

- *blue* indicates nucleotydes which contain adenine or thymine;
- *red* indicates nucleotydes which contain guanine or cytosine;
- *yellow* indicates the overlap of two monomers of the same kind;
- *black* indicates the overlap of two monomers of different kind.

In Fig.1 we show straight strings placed in space; each one is composed of 50 elements and the space contains twenty lines and six columns of these polymeric chains. So, we have 120 chains, that represent a sequence of 6000 nitrogenous bases. The configuration is a magnification of the real phenomena of order of 10^5.

Fig.1 displays two configurations: the left figure is the initial one described above and the right one is its evolution after 250.000 steps. Evolution appears very dynamic and changeable. Some zones empty out, while others become full of entangled or simply near polymers; inside polymers seem to attract or to repel each other in different instants.

This figures turn even more interesting when compared with an image of moving strings segments of the pBR322 DNA of the Escherichia coli (EscoRV-PstI) (937-bp) (Fig.2), kindly given us by Prof. Anita Scipioni of the Chemistry Department of the University of Rome "La Sapienza" [**?**].

We can study our simulation in the following way: consider a vertical projection of the liquid which contains strings and is between two slides; the gravitational field acts on the moving strings by giving a greater probability of shifting southward to the larger molecules. In this projection we do not consider torsion.

Instead, in the images of Figure 2, used for comparison, the field of gravitaty is not considered, because the images come from horizontal projection, but torsion is taken into account. In three dimensions, as torsion as field of gravity would be considered and that induce global dynamics; for this reason it is possible to compose the two dimensional projections.

Fig.3 shows two strings each one with 180 monomers. The top is a *straight* chain and the bottom is a *ladder-like* one. Here, the strings are magnified about 10^5 times.

As before, the left figure displays the initial configurations and the right one the stopped configurations after 100.000 steps. As we can see,

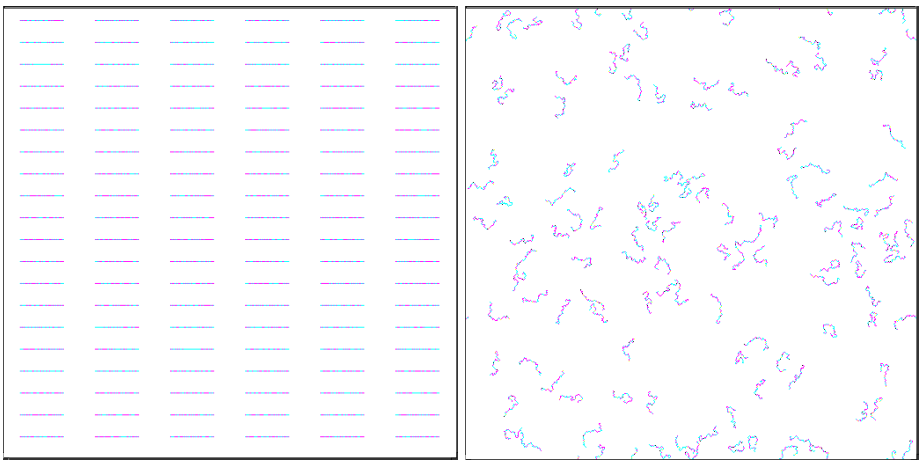

Figure1.

Figure2.

after an initial difference in behaviour, both the strings quickly lose their structure and by looking at the two strings after 100.000 steps we are not able to distinguish between them.

3.2 Closed strings

With at the aim of proving the intrinsic difference between open and closed strings behaviour, we constructed some closed configurations. One example is the *rhomb*, composed of 160 monomers. Every side is placed on a narrow *ladder-like* structure; in fact each monomer is one of the ladder rungs.

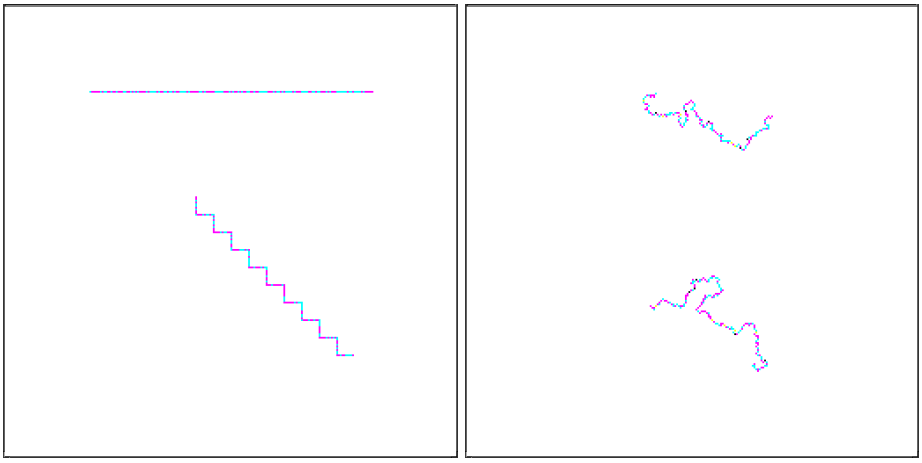

Figure3.

Fig.4 shows the great flexibility of this string. Even if it remains stable around its center of gravity, it is very dynamic and quickly it loses its initial displacement. All the points, beginning with those near the vertices of the *rhomb*, tend to get entangled and spread, alternatively, as if they were about to wrap over each other. But overlapping is forbidden by the "connectivity" and "excluded volume" conditions (see section 4 (1) and (2)), that are closely related to the fact that our simulation is in two dimensions and overlapping for an instant would imply a three-dimensional displacement. So, near strings at any moment go away, at any other spread, and then move together again.

4 Cellular automata and polymer simulations

Because we want to implement our model in a parallel fashion, we use a paradigm of computation, called *cellular automaton*. This device was defined for the first time by John von Neumann and we can describe it as composed by a lattice stucture of identical processors that interact with each other in a strict way, always with the same rule. The single elements are very simple and each interaction involves them all at the same time. So, a cellular automaton can be represented as a regular, infinite, countable lattice of cells, each one being a copy of the same simple

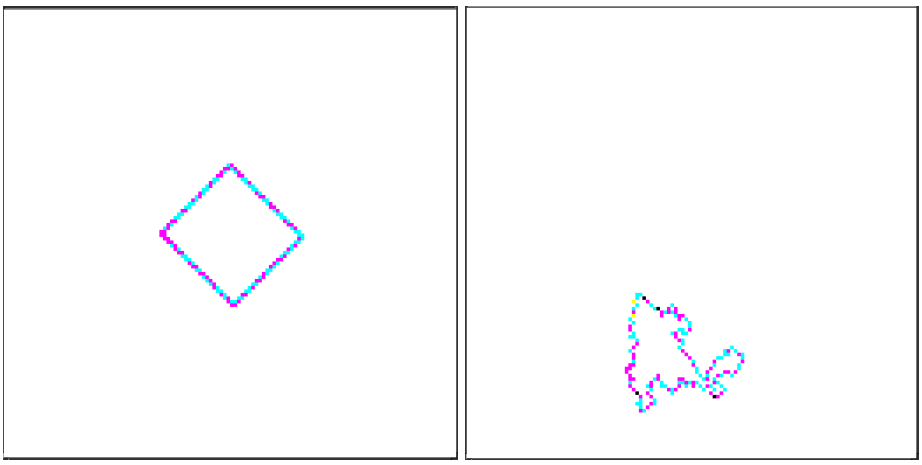

Figure4.

device, placed at the nodes of the lattice and hence evolving at the same time with the same rule. Any cell gets, from a finite number of *near* cells, the internal information and uses it together with its internal state to determine its own evolution.

This paradigm is a very fast model of computation and has a physical implementation in the CAM-8 machine [**?**]. CAM-8 displays the time evolution of the cellular automaton, which corresponds to a succession of configurations; we call the one for $t = 0$ the *initial configuration.*

We briefly describe a method for simulating polymer dynamics. The ball-and-string model first presented in [**?**] and illustrated in (Fig.5). The monomers are rigid balls of radius r_0 and linked to the nearest neighbouring monomers by strings of lenght d_0 with no elasticity. These strings are not allowed to break, as they prevent adjacent monomers from getting

Figure5.

further than d_0 from each other and also the rigid balls are not allowed to overlap.

This first model considers all monomers of the same kind and allows one monomer to move anywhere at distance r_m from its current location

at a time. A single step consists of two parts: first, we select a move; second, we accept or reject the move. The first part consists in a random selection of one monomer and of a direction. The second part states acceptance or rejection according with the following two conditions [**?**]:

(1) *excluded volume*: the chosen monomer does not overlap with another one;
(2) *connectivity*: the move does not make the monomer go further from the neighbors to which it is linked.

We are going to adapt this model, making it become a parallel process implemented by cellular automaton and so we will move one half of monomers at the same time.

5 Our model for DNA chains

We display the strings evolution in a two-dimensional representation, even if by moving one half of monomers at every time step [**?**], we place every chain on two different parallel planes (Fig.6).

Monomers alternate between the planes, so that odd numbered monomers are on one plane and even numbered monomers are on the other one. In this way, the neighbours of each monomer reside in the opposite plane.

Figure6.

We definite a 3×3 region of cells around each monomer in the opposite space as its *bonding neighbourhood*. This is the unique region of cells in which its neighbour reside and no other monomers are allowed to enter. Dynamics are defined by requiring that the motion of every odd or, alternatively, every even monomer be allowed only if its movement to a new position does not add or remove monomers from its bonding neighbourhood. This satisfied the excluded volume (1) and connectivity (2) conditions, avoiding a change in the number of neighbours. This allows one half of the monomers be updated in parallel.

In our model we distinguish among the monomers by means of the

four nitrogenous bases at which nucleotides are bound [**?**]. Differentiation also implies the probability that the direction of movement will differ in each case. We use the molecular characteristics of the monomers for determining the distribution of probabilities. The latter are stated by using the atomic weight and the concentration of two kinds of bases in the real DNA sequence. Random generation routines are designed for this purpose, by respecting the device features, so being local and parallel [**?**] and also suitably tested. Finally, initial configurations are taken from the ADH II (Alcohol dehydrogenase II) of Saccharomyces cervesiae.

The results of our simulations, which we presented in section 3, are very interesting, close to reality and promising for further studies.

6 The algorithm

We place one or more chains on two parallel planes, placing the monomers of the strings in an alternate fashion, so that if we numbered

Figure7.

the monomers along the chain, we would find the even-numbered ones on a plane and the odd ones on the other. This allows us to manage one half of the monomers in parallel (with a multiprocessor like CAM-8) and easily check the connectivity and excluded volume conditions too. In fact, all the *bonding neighbourhood* of every cell lies on the opposite plane. As we want that the movement to respect these rules, exactly two neighbours must be present (except for the extremities) (Fig.7).

In Figure 7 we show an implementation of the "bonding neighbourhood conditions check" for one cell which intends to move eastward [**?**]. Black balls represent monomers on one plane, while white ones represent monomers on the opposite plane. The cells marked by a crossbar must not contain monomers, if we want connectivity and excluded volume conditions be satisfied. In a similar way, the verification can be performed for the other directions.

So, the algorithm can be briefly described as indeterminately repeating the following sequence of steps:

- one plane is randomly chosen;
- every cell of this plane selects a direction (north, south, east, or west) with a probability stated by the kind of monomer contained in it;
- for every monomer, when the direction is stated, the *bonding neighbourhood* conditions are checked;
- for every monomer, if its plane is the chosen one and if its chosen movement is not forbidden, the movement is performed.

Then, the dynamics evolution can be directly displayed on the screen, without any other interface but the choice of colors. So the two planes are linked together, as if we observed them from outside; in this way, the choice of colors can represent differences among monomers.

At last, we can see on the screen a real time evolution, in which all the strings move back and forth on the screen. Indeed, it is possible to fix the evolution images, f.e., every 100 steps and then to show them with a "movie", as done in [**?**]. Here, we chose to present only the initial configuration together with the one obtained after 100.000 steps.

7 The choice of probabilities

We are now going to justify our choice of the distributions of probabilities. With this aim, we designed many original macros in the CAM-8 dedicated language. These programs are able to generate the required distributions, that are different for every cell [**?**] and change with every time step.

The interesting fact is that the generation is performed on-line, without a significant cost for the simulation speed. In fact, suitable expressely coded routines manage a hidden memory of the CAM-8 which, together with its processors, behaves as if it were a slave machine designed for the unique purpose of generating good different distributions of probabilities.

In defining movement probabilities, we observe that the monomers are forced to move by external impulses, surely existing inside a liquid; then these moves can be made only if bonds with neighbour molecules do not forbid it.

7.1 Inertia deriving by hydrogen bonds

Here, our test is to compute the probability that the external impulses push monomers in one direction rather than in another. It is known [**?**]

that the pairs G-C or C-G are more stable than the pairs A-T or T-A; this is probably due to the fact that C-G have three hydrogen bonds, while A-T have only two.

For the random generation, we use 8 bits, that is the numbers rank is [0, .. , 255]. Thus, by supposing that the probability of the A-T move is near 1, we can relate the stability with the number of hydrogen bonds. If we define $2x$ the probability that a monomer with two bonds does not move and we define $3x$ the probability that a monomer with three bonds does not move, we can say that:

$$\begin{cases} 1 - 2x > \dfrac{255}{256} \\ 1 - 3x \leq \dfrac{255}{256} \end{cases}$$

from which we obtain that:

$$\frac{1}{768} \leq x < \frac{1}{512}$$

and so:

$$\frac{1}{256} \leq 3x < \frac{3}{512}$$

Then we can put $x = 1/768$ and $3x = 1/256$.

7.2 The bases weight

At the aim of evalutating the probability of movement of every nucleotide, we want to take into account, atomic weight as well as bonds between bases. Because nucleotides differ from one another for nitrogenous bases, at this time we consider the bases weight only:

Base	*Weight (in daltons)*
A	135.13
T	126.11
G	151.13
C	111.10

As A is bound to T and G to C, we can sum their weight and so we obtain:

$$w_{AT} = w_A + w_T = 261.24$$
$$w_{GC} = w_G + w_C = 262.23$$

As we disposed our strings as if it were between two vertical slides, the probability of going northward or southward is proportional to the mass, i. e. to the atomic weight while probability of going eastward can be assumed to be 1/4.

Instead, probabilities in the north-south directions are stated by the ratio: base weight divided by the sum of it and its complementary one.

Moreover, we do not distinguish between A and T and between G and C, the movement probabilities are conditioned by monomers concentration inside DNA. We cannot compute this percentage in every string, so we use data from genetic biology that establish that percentage in human DNA [**?**] are:

$base$	$percentage\ \ contents$
A	29.3
T	30.0
G	20.7
C	20.0

At last we can assume that the probability that the base is A, supposing that it is A or T, is $\cong 1/2$ and so on.

7.3 Distributions of probabilities

To summerize all our preceding discussions and the results, which we achieved in our simulations, we befine:

$M_0 = \{$a monomer containing A or T moves$\}$
$M_1 = \{$a monomer containing G or C moves$\}$
$B_j = \{$a monomer with B goes j-ward$\}$ (where: $B = A, T, C, G; j = north, south, east, west$)
$p_1 = \mathbb{P}\{$a monomer goes northward$\}$
$p_2 = \mathbb{P}\{$a monomer goes southward$\}$
$p_3 = \mathbb{P}\{$a monomer goes eastward$\}$

$p_4 = \mathbb{P}\{\text{a monomer goes westward}\}$
$p_5 = \mathbb{P}\{\text{a monomer does not move}\}$

and by considering that M_i, $i = 0, 1$, and B_j, $B = A, T, C, G; j = north, south, east, west$, are indipendent events, we obtain the following approximate probabilities:

A: $p_3 = \mathbb{P}(A_e \cap M_0) = \frac{1}{4} * 1 = \frac{1}{4}$

$p_4 = \mathbb{P}(A_w \cap M_0) = \frac{1}{4}$

$p_1 = \mathbb{P}(A_n \cap M_0)\big(1 - \mathbb{P}(A_e \cap M_0) - \mathbb{P}(A_w \cap M_0)\big) = \mathbb{P}(A_n)\mathbb{P}(M_0)\frac{1}{2} =$
$= \frac{126.11}{261.24} * 1 * \frac{1}{2} = \frac{62}{256}$

$p_2 = \mathbb{P}(A_s \cap M_0)\big(1 - \mathbb{P}(A_e \cap M_0) - \mathbb{P}(A_w \cap M_0)\big) = \frac{135.13}{261.24} * \frac{1}{2} = \frac{66}{256}$

$p_5 = 0$

T: $p_3 = \frac{1}{4}$

$p_4 = \frac{1}{4}$

$p_1 = \frac{66}{256}$

$p_2 = \frac{62}{256}$

$p_5 = 0$

G: $p_3 = \mathbb{P}(G_e \cap M_1) = \frac{1}{4} * \frac{1}{256} = \frac{1}{4}$

$p_4 = \mathbb{P}(G_w \cap M_1) = \frac{1}{4}$

$p_1 = \mathbb{P}(G_n \cap M_1)\big(1 - \mathbb{P}(G_e \cap M_1) - \mathbb{P}(G_w \cap M_1)\big) = \mathbb{P}(G_n)\mathbb{P}(M_1)\frac{1}{2} =$
$= \frac{111.10}{262.23} * \frac{255}{256} * \frac{1}{2} = \frac{54}{256}$

$p_2 = \mathbb{P}(G_s \cap M_1)\big(1 - \mathbb{P}(G_e \cap M_1) - \mathbb{P}(G_w \cap M_1)\big) = \frac{151.13}{262.23} * \frac{255}{256} * \frac{1}{2} =$

$$= \frac{73}{256}$$

$$p_5 = \frac{1}{256}$$

C: $p_3 = \frac{1}{4}$

$$p_4 = \frac{1}{4}$$

$$p_1 = \frac{1}{2} * \frac{151.13}{262.23} * \frac{255}{256} = \frac{73}{256}$$

$$p_2 = \frac{1}{2} * \frac{111.10}{262.23} * \frac{255}{256} = \frac{54}{256}$$

$$p_5 = \frac{1}{256}$$

8 Conclusions

We presented a computer simulation system, able to represent the dynamics of DNA segments, when embedded into a non neutral liquid. We described the behaviour of several different displacements, in particular distinguishing between the evolutions of open and closed chains. The chosen representation scale is the molecular one, in which pixels are base (adenine, thymine, guanine, cytosine) molecules and the necessary approximations mainly regard the distinction between adenine or thymine on one side and cytosine or guanine on the other.

Nevertheless, our simulation is close to reality, as shown, for example, in Fig.1 and Fig.2. This is due to the accuracy of the microscopic model as well as the soundness of random generation algorithms, specifically designed for this issue. Possible further improvements are:

(i) distinction among all the bases, not yet completely possible with our model;

(ii) simulation of movement in three dimensions, which requires a new algorithm able to take into account torsion, the one-directional helix and the two dimensional projection of the three-dimensional movement;

(iii) modelization of both the two helixes, possibly with temperature denaturation.

9 Acknowledgements

We want to acknowledge Prof. Ernesto Di Mauro, of the *Genetic and Molecular Biology Department*, University of Rome "La Sapienza", for encouragement, useful discussions, real DNA strings etc. and Dr. Silvio Capobianco, Department of Mathematics, for cooperation, help and kind assistance.

References

1. Astolfi S., *Simulazioni di dinamiche di stringhe di DNA con l'automa cellulare CAM-8*, Tesi di laurea, Dipartimento di Matematica "G. Castelnuovo", Università di Roma "La Sapienza", Anno Accademico 1999-2000
2. Bar-Yam Y., *Dynamics of Complex Sistems (Studies in nonlinearity)*, Reading, Massachussetts, Addison-Wesley, 1997
3. Capobianco S., Mentrasti P., *Generation of on-line randomness by cellular automata*, "Rairo-informatique theorique et applications", 1-16, submitted
4. Di Mauro E., Mentrasti P., Tempestini O., *Open and Closed Polymers Simulations*, CD, Dipartimento di Matematica "G. Castelnuovo", Università di Roma "La Sapienza", 2001
5. Gotoh O., Tagashira Y., *Stabilities of Nearest-Neighbor Doublets in Double-Helical DNA Determined by Fitting Calculated Melting Profiles to Observed Profiles*, "Biopolymers", 20 (1981), 1033-1042
6. Mathews C. K., van Holde K. E., *Biochimica*, Milano, Ambrosiana C. E., 1994
7. Mentrasti P., Clementi A., Pierini P., *L'automa cellulare, un modello di calcolo e di simulazione*, Dipartimento di Matematica "G. Castelnuovo", Università di Roma "La Sapienza", Roma, 1991
8. Mentrasti P., Tempestini O., *Istruzioni fondamentali per l'utilizzo della CAM-8*, Dipartimento di Matematica "G. Castelnuovo", Università di Roma "La Sapienza", Roma, 1999
9. Mishra B., *Alchemy of genoma*, CMU School of Computer Science, 1997
10. Toffoli T., *Programmable Matter Methods, Future Generation Computer System*, Computer System, Domenico Talia and Peter Sloot ed., 1998
11. Toffoli T., Margolus N., *STEP: a Space Time Event Processor, Software Reference*, MIT Laboratory Computer Science, Cambridge, Massachusetts, 1993
12. Vologodskii A. V., Amivikyam B. R., Lyubchenko Y. L., Frank-Kamenetskii M. D., *Allowance for Heterogeneous Stacking in the DNA Helix-Coil Transition Theory*, "Journal of Biomolecular Structure and Dynamics", 2 (1984), 131-147
13. Watson, Hopkins, Roberts, Steitz, Weiner, *Molecular Biology of the gene*, Vol. I, Fourth Editor, USA, 1976
14. Zuccheri G., Scipioni A., Cavaliere A., Gargiulo G., De Santis P., Samori B., *Mapping the intrinsic curvature and flexibility along the DNA chain*, "PNAS", 98 (no. 6) (2001), 3074-3079