

A Probabilistic Model of the DNA Conformational Change

Shiozaki, Masashi

九州大学大学院システム情報科学府：修士2年：理論計算機科学

Ono, Hirotaka

九州大学大学院システム情報科学研究院：助手：理論計算機科学

Sadakane, Kunihiro

九州大学大学院システム情報科学研究院：助教授：理論計算機科学

Yamashita, Masafumi

九州大学大学院システム情報科学研究院：教授：理論計算機科学

<https://hdl.handle.net/2324/25385>

出版情報：2006

バージョン：

権利関係：



A Probabilistic Model of the DNA Conformational Change^{*}

Masashi Shiozaki, Hirotaka Ono, Kunihiko Sadakane, and Masafumi Yamashita

Dept. of Computer Science and Communication Engineering, Kyushu University
{masashio,ono,sada,mak}@tcslab.csce.kyushu-u.ac.jp

Abstract. Predicting the behavior of DNA molecules in vitro is one of the most fundamental issues on DNA computing, but is also known to be quite difficult. Shiozaki et al. proposed a probabilistic model that can simulate many features of biochemical experiments in terms of the reaction rate [7], although there are several differences between the biochemical experiments and the computational simulations on the model. In this paper, we extend the model to support base pairs construction among k DNA sequences, which plays an essential role in realizing branch migrations. The simulation results have much more similarities to the biochemical experiments results than ones on the previous model, which implies that the analysis of the model may give some insight about the reaction rate. Through the analysis, we conclude this paper by giving a guideline for designing DNA sequences that can quickly react.

1 Introduction

Predicting the behavior of DNA molecules in vitro is one of the most fundamental issues on DNA computing, but also known to be quite difficult. Many researchers try to understand the principle of DNA molecule reactions from several viewpoints. The aim of this paper is to understand the mechanism of the DNA conformational change.

Over the past few decades, modelling the behavior of RNA (not DNA) has been intensively investigated [1, 2, 8, 10]. Most of the models are essentially based on Markov process under the assumption that an RNA sequence changes its own structure according to the transition probability defined by the free energy. Some models succeeded; they can simulate the folding behavior of RNA well. These models are based on the simplicity of the RNA reaction; a base pair of RNA is made only in one sequence¹. On the other hand, the DNA reactions are usually more complicated, because base pairs of DNA are made not only on one sequence but also on two or more sequences. That is, the behavior of DNA molecules is

^{*} This research partly received financial support from Scientific research fund of Ministry of Education, Culture, Sports, Science and Technology.

¹ This is actually a bit exaggerating and in RNA base pairs among more than one sequence are made. What we want to say here is that in RNA reaction a base pair construction in one sequence is known to be essential.

supposed to be quite different from that of RNA, and the same technique is not applicable directly. In fact, there had been no study on the analysis of the DNA behavior, as far as the authors know.

Among them, Shiozaki et al., the authors of this paper, proposed a probabilistic model based on Markov process for simulating the behavior of DNA, as a first step of the DNA behavior's analysis [7]. Considering the hybridization of multiple DNA sequences, we introduced a model with two macro states and the coupling factor p that can control the state transition between the two macro states. The model can treat the structures formed with one or two sequences, which are not so essential in the RNA conformation but are crucial in the DNA conformation. We conducted computational simulations to compare with biochemical experiments, whose results are provided by Prof. Suyama's Group ². That biochemical experiments observe the reaction rate in which two types of DNA sequences (one type of sequences and its complement type of sequences) hybridize in vitro under a constant temperature. Through the comparison studies, we see that our computational simulations could imitate the biochemical experiments on some of their typical features, though some features are still left to be realized.

In this paper, we propose an advanced model, which is an extension of the previous model, so as to explain the unrealized features. The previous model considers the structures formed with only at most two sequences. This implies that some conformational changes in which more than two sequences react, such as branch migrations, are not supported in the model. It is a problem, because branch migrations are considered to play a very important role in DNA reactions. A point of the new model is that it permits any DNA sequence to have base pairs among k sequences, which make it possible to simulate the mutual influences of multiple DNA sequences. For example, a double stranded structure makes a base pair with other structures; the branch migrations also may occur in the model.

We then conduct computational simulations on the new model. Simulation results show more similarities to biochemical experiments than the previous one. Through the computational simulations, we propose a guideline for designing DNA sequences that quickly react, and further simulation studies support the availability of the guideline.

These simulation studies on our model correspond to the biochemical experiments explained above. Simulation studies for another kind of DNA reaction (biochemical experiments) in Prof. Hagiya's group ³, which is about the branch migrations, also show that they can nicely imitate the bioexperimental results. The similarities between the simulation results on our model and the biochemical experiments imply that our model may be a good approximation of the DNA conformational system in vitro; if our model approximates essential parts of DNA conformation indeed, we can estimate the DNA reactions without any biochemical experiments, which means that we can save many costs of biochemical experiments, i.e., a lot of money, time, labor, and so on. Of course, more

² <http://dna.c.u-tokyo.ac.jp/>

³ <http://hagi.is.s.u-tokyo.ac.jp/>

computational simulations and more biochemical experiments should be performed to see the validity of the model, if yes, our results may be a giant leap for DNA computing.

2 Preliminaries

In this section, we first give a basic simulation model of RNA conformation. We then explain the DNA reactions for simulating the DNA conformational change and show the results of the biochemical experiments.

2.1 The simulation model of RNA

Wolfinger et al. propose a model for simulating the folding behavior of RNA, which is based on Markov process [10]. In their study, each state in the Markov process corresponds to each RNA structure. They assume that RNA sequences change their own structures according to the state transition probability defined by the free energy of each structure. In their simulation model, the neighborhood of the state is defined as the set of structures that are formed from the present structure with a single base pair variation. The state transition probability is as follows:

$$P(X_{t+1} = y | X_t = x) = \frac{r_{xy}}{\sum_{z \in N(x)} r_{xz}}, \quad (1)$$

where $N(x)$ is the neighborhood of the state x and r_{xy} is the transition rate from the state x to y defined as follows:

$$r_{xy} = \exp \left\{ -\frac{E_{yx}^\# - E(x)}{RT} \right\}, \quad E_{yx}^\# = \max\{E(x), E(y)\}.$$

The transition state energies $E_{yx}^\#$ assures the detailed balance, $E_{yx}^\# = E_{xy}^\#$. As a result, this model could imitate the folding behavior of RNA quite well. The model is based on the independence of the past history and the random behavior of RNA conformational change. Since these also hold for DNA conformation, we expect that the behavior of DNA can be modelled as a Markov process, but straightforward application is not possible because in the case of DNA reactions among multiple sequences may occur easily; it is difficult to define “structures” and their free energy.

2.2 Basic DNA reactions

We consider the situation that one type of DNA sequences (we call them primary sequences) and the same number of its complements are put in vitro at the same time under a constant temperature. At the initial state all DNA sequences have no base pair. Then DNA sequences start to change their own structures randomly and after a while they form the perfect double stranded structures.

This is considered one of the most basic DNA reactions, and actually Suyama et.al. also concentrate to analyze this basic situation.

In the biochemical experiments conducted by Suyama et al., they use six types of DNA sequences that have the same length and almost the same minimum free energy (MFE) of the structure that is formed with a primary sequence and its complement. Table 1 presents the primary sequences of six instances and the MFEs that are calculated with a Vienna-RNA1.4 package [3, 4] which was yet improved by Uejima [9]. In this table, MFEs are minimum relative free energies with respect to the structure with no base pair.

These chemical experiments were done in the following environment; the absolute temperature is 298 K and the concentration of DNA is 10 nmol/l. For comparison, we use three instances #60, #171 and #176, each of which has obvious characteristics in terms of the reaction rate.

Table 1. The primary sequences and MFEs of the six instances.

ID	Primary Sequence	MFE (kcal/mol)
60	TTCGCTGATTGTAGTGTGCACA	-36.16
171	CGCGATTCCTATTGATTGATCCC	-34.78
176	GGGATCAATCAATAGGAATCGCG	-37.10
117	CGCGATTCCTATTGATTGATCCC	-35.60
227	TAGCACCCGTTAAAACGGAAATG	-35.27
247	GCCTCCTTAGTAAGAACTGGTGC	-36.14

Figure 1 shows the experimental results in biochemistry⁴ of #60, #171 and #176. The horizontal and vertical axes represent real time and fluorescence intensity, respectively. In these experiments, fluorescence intensity is proportional to the number of stacked base pairs between two sequences.

Figure 2 shows the simulation result of Shiozaki et al. [7]. The horizontal and vertical axes represent the number of steps and the sum of the number of stacked base pairs, respectively. We can see that the biochemical experiments and the simulations have some common tendencies. For example, both of the figures show that the reactions of #176 are slow. This is because #176 tends to make base pairs in one sequence frequently. Table 2 presents the MFEs in one of the primary sequences and the complements of #60, #171 and #176. It is known that the lower the free energy is, the more stable the structure is. Hence that means both of the primary sequence of #176 and its complement can form hairpin structures whose free energy are lower than the structure with no base pair, while those of #60 and #171 do not form such structures in one sequence. Figure 3 shows the structures corresponding to the MFEs in Table 2. Forming a hairpin structure prevents a strand from hybridizing with another to make a

⁴ This figure was offered by Prof. Suyama of the University of Tokyo.

duplex. This may be the reason that the reaction of #176 is much slower than the other two.

These are the features of the biochemical experiments that our simulation can imitate. On the other hand, there are some differences between the results of our basic model and the biochemical experiments. For example, the results of the biochemical experiments show that the DNA sequences of #60 react faster than those of #171, while our simulation model shows the opposite results.

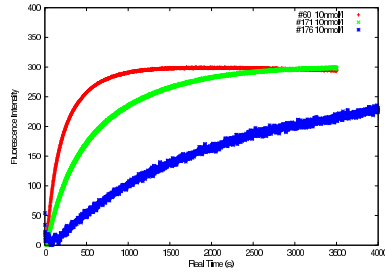


Fig. 1. Experimental results

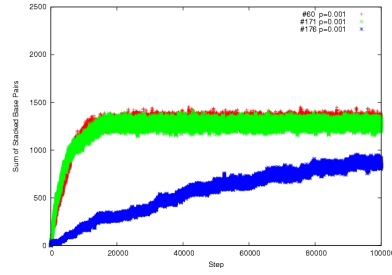


Fig. 2. Simulation results

Table 2. MFE in one sequence of #60, #171 and #176

MFE in one sequence (kcal/mol)		
ID	Primary Seq	Complement
60	0.0	0.0
171	0.0	0.0
176	-4.92	-3.88

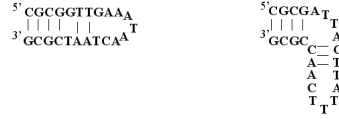


Fig. 3. Structures of the MFE in one sequence of the primary sequence of #176 (left) and its complement (right)

2.3 Branch migrations

In the previous subsection, we explain a basic DNA reaction, its biochemical experiments and the simulation results of our old model. In this section, we introduce two new reactions about branch migration, whose biochemical experiments are provided by Prof. Hagiya group. We call them BM1 and BM2, respectively.

BM1 is described in Figure 4. We prepare three types of DNA sequences; these are sequences consisting of 40 bases (normal color one in the figure), called the primary sequences, their complement sequences (shaded one) and the half sequences of the complement (normal color one but the length is half). A primary

sequence has a black hole quencher (BHQ) at the 3' end of the sequence. On the other hand, a half of its complement has a fluorescent material, FAM at the 5' end of the sequence. At the initial state, these two types of sequences form double stranded DNA structures (Figure 4 (1)); that is, BHQs neighbors FAMs and the light of FAMs is absorbed by BHQs. Therefore the molecules do not glow at this moment. As time advances, shaded sequences start to make base pairs with the primary sequences (Figure 4 (2)). Then branch migrations occur and the half of their complements with a FAM start to dissociate from the primary sequence with a BHQ. Finally the primary sequences and their complements form perfect double stranded structures (Figure 4 (3)).

BM2 is similar to BM1, but four types of sequences participate. Figure 5 explains BM2. We prepare three types of DNA sequences as BM1 and another type of sequences; We add the half sequences of the primary sequence, that is the length is 20. These sequences also make double stranded DNA structure with the complements of the primary sequences at the initial state (Figure 5 (1)). Hence there exist two kinds of double stranded DNA structures. If the reactions perfectly occur, then two types of perfect double stranded structures are formed in the equilibrium state (Figure 5 (3)).

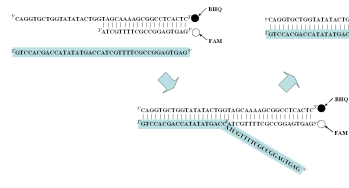


Fig. 4. The reaction of BM1 (1,2,3)

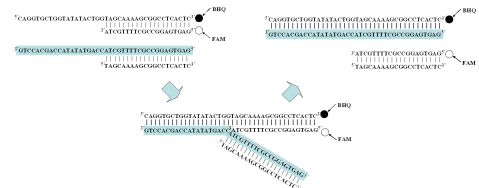


Fig. 5. The reaction of BM2 (1,2,3)

In these two experiments, molecules glow when BHQs and FAMs stand off to each other, which means that fluorescence intensity becomes strong after the branch migrations completely finishes. Hence the fluorescence intensity shows the progress of the reaction.

Figure 6 shows the results of the biochemical experiments ⁵. The horizontal and the vertical axes represent real time and fluorescence intensity respectively. These experiments are done in the environment where the absolute temperature is 318 K and the concentration of DNA is 100 nmol/l.

As shown in this figure, BM1 reacts in a very short time while BM2 takes a lot of time to finish branch migrations completely.

⁵ This result is offered by Prof. Hagiya of the University of Tokyo.

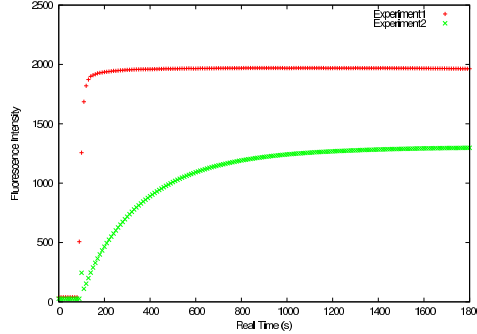


Fig. 6. Results of biochemical experiments

3 New Simulation Model

As mentioned in the last paragraph of Section 2.2, old results [7] have some differences from those of the biochemical experiments in terms of the reaction rate of #60 and #171. That is, the old model may extract some part of the DNA reaction, but that extraction is rough and not good enough. In this section, we propose a more polished model, which can simulate making base pairs among k DNA sequences, and then show the simulation results.

3.1 The simulation model with the extended neighborhood

This model considers any structure created by k DNA sequences, so that the branch migration may be realized. Although the branch migration is considered essential for DNA reactions, the old model does not support; if it is the reason that the old model is not good enough, we may expect good improvements for the new model.

In the design of Markov process system as a simulation model, there are several computational problems to be considered, because our objective, DNA conformation, has an enormous number of states (DNA structures). The first problem is how to define the transition. Given a state x (DNA structure), We call a set of states (DNA structures) that can be changed from x , *neighborhood* of x . To simulate a Markov process, we compute all the transition probability from x to the neighborhood. That is, it is necessary to enumerate the neighborhood of x . However, enumerating all the neighborhood is too time-consuming if we adopt very general models of DNA reaction. On the other hand, it is considered that the transitions from x to most of neighbor states of x in such a model seldom occur. By these observation, we consider a model that has a restriction where k DNA sequences are not affected by any other sequences and all of them are neighbored at any time. In our model, we define the neighborhood as the set of

structures that is formed with k DNA sequences from the present structure with a single base pair variation; it includes the present structure. By this, the size of the neighborhood is $O(k^2 l^2)$ at any time.

Preparing k DNA sequences, we move to the next state selected from all the neighborhood according to the state transition probability at each step; we use the same transition probability with [7] which is defined as follows:

$$P(X_{t+1} = y | X_t = x) = \frac{R_{xy}}{\sum_{z \in N(x)} R_{xz}}, \quad (2)$$

where $N(x)$ is the neighborhood of a state x and R_{xy} is the transition rate defined as follows:

$$R_{xy} = \begin{cases} p \cdot r_{xy} & \text{(a first base pair between two sequences)} \\ r_{xy} & \text{(otherwise),} \end{cases}$$

However, enumerating all the neighborhood in this model is still too complicated when k is large. This is because we need to change the order of the k sequences to get the free energy of each structure in the neighborhood. We implement this model as $k = 3$ and $k = 4$ with computer so far.

In order to simulate the basic DNA reactions we use this model as $k = 4$. Preparing two primary sequences and the same number of their complements, we can simulate any pairs of them. In order to realize branch migrations, we use this model as $k = 3$ and $k = 4$. In fact, in the BM1 and BM2, three and four types of DNA sequences participate in the reaction.

3.2 Computer Simulations of Basic DNA Reactions

We use three types of DNA sequences (#60, #171 and #176) in Table 1 as instances and do Monte Carlo simulations 50 times. Then the total number of sequences is 200, which is the same to the previous simulation [7]. Considering the results of the previous simulation model, we set $RT = 2.0$ and $p = 0.0001$. These parameters seem to depend on the concentrations.

Figure 7 shows the results of 0 to 10^6 steps when $p = 0.0001$. The horizontal and vertical axes represent the number of steps and the sum of the number of stacked base pairs of 50 times Monte Carlo simulations. In this figure, the rank of the reaction rate of #60, #171 and #176 is the same as the results of biochemical experiments, while not in the previous simulations. Therefore this model can simulate the DNA reactions well.

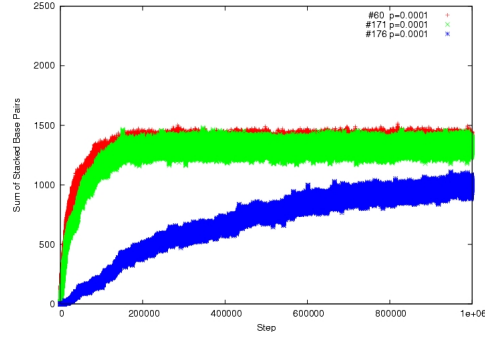


Fig. 7. The results of our advanced simulation model ($p = 0.0001$)

Discussion It is natural to consider that these results come from the realization of the branch migration. Intuitively if the MFEs with two primary sequences and with their complements are small, the reaction becomes slow. In fact these MFEs of #60 are smaller than those of #171. Therefore, in terms of the free energy, #171 is supposed to react faster than #60. However the results of the biochemical experiments are the opposite to this prediction. This fact implies that we can not judge what DNA sequences react fast only from the free energies.

We then focus on the structures of the MFEs with two primary sequences and their complements. Table 3 represents these MFEs and the number of bases of *free 5' end* and *free 3' end* of #60, #171 and #176; the free 5' (3') end is the number of bases that form no base pair at the 5' (3') end of the structure. Since two DNA sequences form these structures, we show two numbers of bases of free 3' end and 5' end for each. The more the number of bases of free 5' (3') end is, the more easily the other DNA sequences can make base pairs with the structure. Especially if both of free 5' (3') end of two primary sequences and that of free 3' (5') end of their complements are large, the branch migration occurs with ease and the reaction can be accelerated.

As presented in Table 3, the structure of two primary sequences of #60 and their complements have many free bases that do not make base pairs at the 5' and 3' end respectively; it implies that the branch migration occurs with ease when the DNA sequences form such structures. On the other hand, that of #171 and its complement have less free bases at both of the edges. We consider that this is the reason that #60 reacts faster than #171 in biochemical experiments.

These discussions also explain the slow reaction of #176. As shown in Table 3, #176 has much lower MFEs. In addition, both of these structures have no free base that does not make base pairs at the edges. Therefore the branch migration hardly occurs and the reaction is not accelerated.

Table 3. The characteristics of the structures of MFEs with two primary sequences and their complements of #60, #171 and #176

ID	Primary Seq			Complement		
	MFE(kcal/mol)	free 5' end	free 3' end	MFE(kcal/mol)	free 5' end	free 3' end
60	-8.54	13, 13	1, 1	-6.60	0, 1	13, 14
171	-4.98	3, 3	7, 7	-6.22	7, 7	3, 3
176	-14.59	0, 0	0, 0	-10.42	0, 0	0, 0

Now we consider the reaction rate of six DNA sequences including #117, #227 and #247. Table 4 presents the MFEs with two primary sequences and their compliments and the number of bases of free 5' end and free 3' end of #117, #227 and #247. From the above discussions and Tables 3 and 4, we predict the following:

- The reactions of #117 and #60 are much faster than the others.
- The reactions of #171 and #227 are a bit slower.
- The reaction of #176 is the slowest.

Table 5 shows the rank of the reaction rate of the six types of DNA sequences in the biochemical experiments for each concentration. The predictions seem to explain the results of the biochemical experiments especially when the concentration of DNA is high. The higher the concentration is, the more frequently the DNA sequences encounter with each other; it means that the branch migrations easily occur. Therefore these predictions match the experimental results better when the concentration is high.

Table 4. The characteristics of the structures of MFE with two primary sequences and their complements of #117, #227 and #247

ID	Primary Seq			Complement		
	MFE(kcal/mol)	free 5' end	free 3' end	MFE(kcal/mol)	free 5' end	free 3' end
117	-8.49	0, 0	19, 19	-8.25	19, 19	0, 0
227	-14.84	6, 6	5, 5	-14.62	5, 5	6, 6
247	-9.27	5, 5	9, 9	-6.85	0, 17	8, 0

There still remain some problems. The comparison between the results of the biochemical experiments and our simulation model including #117, #227 and #247 shows some differences. These may be due to the lack of the dynamical elements. It is known that DNA sequences start to make base pairs from an endpoint and they sequentially bond like zippers. Nevertheless our model can not simulate such reactions.

We conclude this section by the following.

1. We have almost succeeded in simulating the DNA reactions by the probabilistic model based on the free energy of each structure.
2. However, only the free energy information may not be enough to simulate the precise behavior of DNA.
3. A new model reflecting the dynamical elements might approximate DNA reactions much better.

Table 5. Ranking of strands by reaction rates in the biochemical experiments

	Concentration		
Rank	5 nmol/l	10 nmol/l	20 nmol/l
1	60	247	117
2	247	60	60
3	117	117	247
4	227	227	227
5	171	171	171
6	-	176	176

3.3 Computer Simulations of Branch Migrations

We apply our advanced model as $k = 3$ and $k = 4$ to simulate BM1 and BM2, respectively. We set $RT = 1.0$ and do Monte Carlo simulations 200 times.

Figure 8 shows the results. The horizontal and vertical axes represent the number of steps and the number of structures in that a BHQ and a FAM are away from each other (we call them the fluorescent structures), respectively. We can see many common tendencies between the simulation results and the biochemical experiments also in these reactions; for example, BM1 reacts in a very short time while BM2 takes a lot of time to be in the steady state. These results support that our model is useful not only for the basic DNA reactions but also for more complicated DNA reactions.

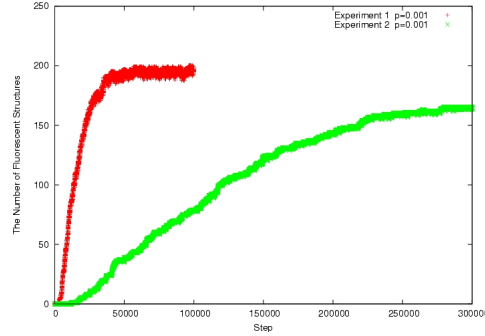


Fig. 8. The comparison between BM1 and BM2 ($p = 0.001$).

4 Conclusion

In this paper, we propose a new probabilistic model of DNA conformational change. The simulation studies for two basic DNA reactions show that the model may be a good approximation of (actual) biochemical reaction. Although further improvements may be possible, the authors think our model has succeeded in extracting some essence of the DNA reactions. To confirm this, much more simulation and biochemical experiments are needed. Also some (computational) theoretical studies might be useful and interesting to analyze thermodynamical properties of DNA reactions.

As a byproduct of the simulation studies, we obtain a guideline for designing DNA sequences that can react fast in terms of the basic DNA reactions; these are summarized as follows:

1. Both the primary sequence and its complement should not form hairpin structures (from the results of [7]).
2. Both two primary sequences and their complements also should not form the stable structure; in the case that they form such structures, the numbers of bases of free 3' end and 5' end of those should be large (Subsection 3.2).

Acknowledgments

The authors sincerely appreciate Professor Suyama and Professor Hagiya of the university of Tokyo who offered experimental data in biochemistry to us. We also thank the anonymous reviewers for their insightful comments to improve the presentation of the paper.

References

1. C. Flamm, *Kinetic Folding of RNA*, Dissertation, 1998.

2. C. Flamm, W. Fontana, I. L. Hofacker and P. Schuster, *RNA Folding at Elementary Step Resolution*, Santa Fe Institute 12-078, 1999.
3. I. L. Hofacker, *The Vienna RNA Secondary Structure Server*, Nucleic. Acids Res. 31 3429-31, 2003.
4. I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer M. Tacker and P. Schuster, *Fast Folding and comparison of RNA Secondary Structured*, Monatsh. Chem. 125 167-88, 2003.
5. A. J. Hartemink and D. K. Fifford, *Thermodynamic Simulation of Deoxyligonucleotide Hybridization for DNA Computation*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science.
6. S. Kobayashi, *Strand Design for Molecular Computation*, Sciences, 2003.
7. M. Shiozaki, H. Ono, K. Sadakane and M. Yamashita, *Modeling DNA Conformation Change and Theoretical Analysis on the Reaction Rate*, Preproceedings of the 11th International Meeting on DNA Computing (DNA11), 408, 2005.
8. M. Tacker, W. Fontana, P. F. Stadler and P. Schuster, *Statistics of RNA Melting Kinetics*, Eur. Biophys. J., 23:29-38, 1993.
9. H. Uejima, M. Hagiya, *Analyzing Secondary Structure Transition Paths of DNA/RNA molecules*, PreProceedings of the 9th International Meeting on DNA Computing (DNA9) 4731-4741, 2004.
10. M. T. Wolfinger, W. A. Svrcek-Seiler, C. Flamm, I. L. Hofacker and P. F. Stadler, *Efficient Computation of RNA Folding Dynamics*, J. Phys. A: Math. Gen. 37: 4731-4741, 2004.
11. M. Zucker, *The Equilibrium Partition Function*, <http://www.bioinfo.rpi.edu/zukerm/lectures/RNAfold-html/node3.html>, 2003.