Dinucleotide Step Parameterization of Pre-miRNAs Using Multi-objective Evolutionary Algorithms

Jin-Wu Nam[1,2], In-Hee Lee[3], Kyu-Baek Hwang[4],
Seong-Bae Park[5], and Byoung-Tak Zhang[1,2,3]

[1] Graduate Program in Bioinformatics
[2] Center for Bioinformation Technology
[3] Biointelligence Laboratory, School of Computer Science and Engineering
Seoul National University, Seoul 151-742, Korea
[4] School of Computing, Soongsil University, Seoul 156-743, Korea
[5] Department of Computer Engineering
Kyungpook National University, Daegu 702-701, Korea
btzhang@bi.snu.ac.kr

Abstract. MicroRNAs (miRNAs) form a large functional family of small noncoding RNAs and play an important role as posttranscriptional regulators, by repressing the translation of mRNAs. Recently, the processing mechanism of miRNAs has been reported to involve Drosha/DGCR8 complex and Dicer, however, the exact mechanism and molecular principle are still unknown. We thus have tried to understand the related phenomena in terms of the tertiary structure of pre-miRNA. Unfortunately, the tertiary structure of RNA double helix has not been studied sufficiently compared to that of DNA double helix. The tertiary structure of pre-miRNA double helix is determined by 15 types of dinucleotide step (d-step) parameters for three classes of angles, i.e., twist, roll, and tilt. In this study, we estimate the 45 d-step parameters (15 types by 3 classes) using an evolutionary algorithm, under several assumptions inferred from the literature. Considering the trade-off among the four objective functions in our study, we deployed a multi-objective evolutionary algorithm, NSGA-II, to the search for a nondominant set of parameters. The performance of our method was evaluated on a separate test dataset. Our study provides a novel approach to understanding the processing mechanism of pre-miRNAs with respect to their tertiary structure and would be helpful for developing a comprehensible prediction method for pre-miRNA and mature miRNA structures.

1 Introduction

The tertiary structure of RNAs is deeply related with their processing and functions, and knowing it helps to resolve their binding mechanisms with other molecules in cells. It can be described by several angle parameters, such as twist, roll, and tilt. However, the tertiary structure of RNA has not been studied much yet, compared to the secondary structure [1] or the tertiary structure of DNA [2]. Crystallography methods facilitated the elucidation of the structural parameters of DNA double helix [2, 3].

Recently, the known functional range of RNA has been expanded gradually since a new posttranscriptional regulator, (i.e., the microRNA (miRNA)), was found [10].

E. Marchiori, J.H. Moore, and J.C. Rajapakse (Eds.): EvoBIO 2007, LNCS 4447, pp. 176 – 186, 2007.

miRNAs are defined as single-stranded RNAs of ~22 nucleotides (nt) in length, generated from endogenous transcripts that can form local hairpin structures [11]. The local hairpin structures are processed by the nuclear RNase type III enzyme, Drosha, releasing the hairpin-shaped intermediates (pre-miRNAs) which are typically 60-70 nt [12]. After exported to the cytoplasm, the pre-miRNAs are cleaved by another RNase III type enzyme Dicer and then are processed into the miRNAs [13]. However, the structural mechanism associated to the recognition and processing of pre-miRNA still remains unknown. Elucidation of the structural mechanism is one of the most crucial problems towards the understanding of molecular basis of miRNA processing. In a recent report on the structural mechanism, it is shown that the cleavage site by Drosha is distant from a terminal loop by about two-turn helices[14]. This biological knowledge can be used for the parameterization of tertiary structure of pre-miRNA.

The parameterization of tertiary structure is a computationally intensive work involving a large number of parameters and several objectives. Conventional approaches to the parameterization include iterative algorithms and weighted linear sum methods. Iterative methods optimize each of the objectives one by one until a self-consistent state has been reached [4]. Weighted linear sum methods address the multiple objective problem by choosing suitable weighting parameters between objectives [5]. However, the choice of weighting parameters for different objectives becomes another challenging problem when the different objectives are dependent on each other. In this case, *a priori* knowledge about the dependency structure is needed for finding suitable weight values.

An alternative method is to adopt a multi-objective optimization algorithm where all objectives are simultaneously optimized. Multi-objective optimization algorithms have widely been applied to the problems where the trade-off relation (dependency) among the objectives exists [6-8]. They produce not a single parameter set but a variety of parameter sets with various trade-offs for the objective functions. Multi-objective optimization algorithms find the optimal solution by comparing the candidate solutions based on the dominance relationship. When comparing two solutions with respect to the dominance relationship, the fitness value of each objective is considered together. Therefore, there is no information distortion in the multi-objective optimization algorithm, whereas the weighted linear sum method inevitably distorts some information while summarizing the individual fitness values [9].

We introduce a novel approach for the parameterization of pre-miRNA structure using multi-objective evolutionary algorithms (MOEAs). In our knowledge, there has been no reported research on the parameterization of pre-miRNAs in terms of their tertiary structure. In specific, we focus on the dinucleotide step (d-step) parameters of double helix structure of pre-miRNA. Results of this study may help to understand the mechanism as well as can be used as an integral part for the prediction of pre-miRNAs. This paper largely consists of three parts; the first section describes the implementation of MOEAs for the parameterization of RNA tertiary structure; the second part demonstrates the results of a case study about pre-miRNAs; discussion and future work are given in the last part.

2 Materials and Methods

2.1 Tertiary Structure of RNA

The double helix structure of RNA is determined by three angle parameters; twist (*W*), roll (*R*), and tilt (*T*) as described in Figure 1. Twist is a main angle of the helix structure and decides whether it is left-handed or right-handed (Figure 1(b)). Roll is a rotation angle bending along the main groove and a minor groove, and it opens the grooves (Figure 1(c)). Tilt is a rotation angle where a plane of base pair moves up and down, and it destabilizes the stacking energy (Figure 1(d)). The angle parameters determine the position of each base and backbone in three dimensional spaces. They are basically

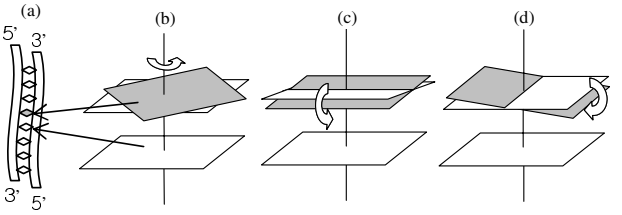

Fig. 1. The angle parameters of RNA tertiary structure. (a) The double helix structure of RNA; (b) Twist; (c) Roll; (d) Tilt. The rhombus denotes the plane of base pair.

defined with dinucleotide step (d-step) parameters, as described in the following. Here, we can define three net angles as described in Equations (1)~(3) [15]. Net twist ($\tilde{W}_j$) is the cumulative twist from the first to the last d-step (*L* d-steps) of the *j*th example. Net roll ($\tilde{R}_j$) and net tilt ($\tilde{T}_j$) are the cumulative angles from the first to the last d-step of the *j*th example.

$$\tilde{W}_j = \sum_{i=1}^{L} W_i \tag{1}$$

$$\tilde{R}_j = \sum_{i=1}^{L}\left[-T_i \cdot \sin\left(\sum_{k=1}^{i} W_k \right) + R_i \cdot \cos\left(\sum_{k=1}^{i} W_k \right) \right] \tag{2}$$

$$\tilde{T}_j = \sum_{i=1}^{L}\left[T_i \cdot \cos\left(\sum_{k=1}^{i} W_k \right) + R_i \cdot \sin\left(\sum_{k=1}^{i} W_k \right) \right], \tag{3}$$

where W_i, T_i, and R_i are d-step parameters for twist, tilt, and roll angles, respectively.

2.2 RNA Dinucleotide Step

The secondary structure of a single RNA sequence can be predicted using the nearest neighbor model [16]. It considers thermodynamic interaction between base pair *i* and *i*+1, called dinucleotide step (d-step). Here, each d-step has three tertiary structural

parameters: twist, roll, and tilt angles. The d-steps can be classified into the three types according to bases' tendency or effect over geometric space. In this study, we use the following ten d-steps.

- Pyrimidine(Py) – Purin(Pr) steps: UA(UA), CA(UG), CG(CG)
- Pr – Py steps: AA(UU), AG(CU), GA(UC), GG(CC)
- Pr – Pr steps: AU(AU), GU(AC), GC(GC)

Here, the complement pair of a dinucleotide step is given in parentheses. Also, we need the following five additional d-steps for insertion or deletion. They represent an internal bulge in RNA structure.

- -A(U-), -G(C-), -U(A-), -C(G-), --

Here, '-' denotes deletion. To summarize, we have to consider 15 d-steps for each angle parameter (twist, roll, and tilt) and thus we should search for the 45 d-steps parameter values (15 d-steps by 3 angle classes).

2.3 Description of the Objective Functions

Here we present the four objective functions used to search for the optimal d-step parameters of double helix structure of pre-miRNA. For the optimization of the d-step parameters, we exploit some prior knowledge, which is specific to the pre-miRNAs. Since the same preprocessing mechanism is applied to all pre-miRNAs, their tertiary structures should be similar. The first function restricts the net twist should be similar across all miRNAs and the second implies the sense-antisense pair should have similar twist values. The third and fourth denotes that the net roll values and the net tilt values should be similar for all examples. All objective functions should be minimized here.

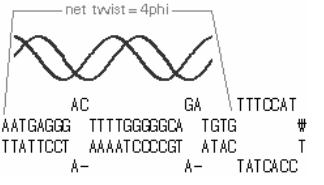

Fig. 2. The net twist of pre-miRNA

2.3.1 Twist: Mean of Difference and Standard Error

All net twists from a Drosha cleavage site to a Dicer cleavage site should be similar to be cropped by the same processing mechanism. A previous experiment reported that the net twist is about two-turn helices (4π) as depicted in Figure 2 [14]. Based on this, we can define the first objective function, which minimizes the difference from 4π and the standard errors of the net twists. If there are N pre-miRNAs, the mean of difference (MOD) and the standard error of the net twists are defined as:

$$MOD = \frac{\sum_{j=1}^{N} \left| 4\pi - \tilde{W}_j^c \right|}{N} \quad (4) \qquad SE_W = \sqrt{\frac{\sum_{j=1}^{N} \left(E(\tilde{W}_j^c) - \tilde{W}_j^c \right)^2}{N}}, \quad (5)$$

where $\tilde{W}_j^c$ is the net twist of the *j*th pre-miRNA calculated by a set of d-step parameters, *C*. From Equations 4 and 5, the first objective function can be defined as follow:

$$f_1 = MOD + SE_W \tag{6}$$

Here, SE_W was added because the net twist should be uniformly distributed around the two-turn helices.

2.3.2 Twist: Mean of Difference Between Both Strands

The stem structure of RNA is bi-directional: (1) from the stem-end of the mature miRNA to the loop-end of the mature miRNA and (2) vice versa. Thus, there are two net twists for a sense and an antisense strand, on a stem structure of pre-miRNAs. Here, a tuple $(S_j^C, \overline{S}_j^C)$ specifies the net twists of the sense and antisense strands of the *j*th pre-miRNA calculated by a set of d-step parameters, *C*. We can assume that there is no difference of the net twists between both strands. Hence, we can define the second objective function f_2 using the mean of difference between both strands (MDS).

$$f_2 = MDS = \frac{\sum_{j=1}^{N}\left(\left|S_j^C - \overline{S}_j^C\right|\right)}{N} \tag{7}$$

2.3.3 Roll: Standard Error of Net Rolls

The third objective function is designed to minimize the standard error of net rolls. The standard error of net rolls is defined as follow:

$$f_3 = SE_R = \sqrt{\frac{\sum_{j=1}^{N}(E(\tilde{R}_j^c) - \tilde{R}_j^c)^2}{N}} \tag{8}$$

where, $\tilde{R}_j^c$ is the net roll of *j*th pre-miRNA calculated by a set of d-step parameters, *C*.

2.3.4 Tilt: Standard Error of Net Tilts

The last objective function minimizes the standard error of net tilts. The standard error of net tilts is defined as follow:

$$f_4 = SE_T = \sqrt{\frac{\sum_{j=1}^{N}(E(\tilde{T}_j^c) - \tilde{T}_j^c)^2}{N}}\ , \tag{9}$$

where $\tilde{T}_j^c$ is the net roll of the *j*th pre-miRNA calculated by a set of d-step parameters, *C*.

2.4 Multi-objective Optimization

A multi-objective optimization is to find a set of decision vectors $\vec{x}^*$ which minimize (or maximize) *m* objective vectors (functions) $\vec{f}(\vec{x})$ at the same time.

$$\vec{x}^* = \arg\min_{\vec{x}}(\vec{f}(\vec{x})) \tag{10}$$

$$\vec{f}(\vec{x}) = \left(f_1(\vec{x}), f_2(\vec{x}),..., f_m(\vec{x})\right), \tag{11}$$

where $f_i(\vec{x})$s are objective functions. The decision vector (parameters) $\vec{x}$ consists of n real values $(x_1, x_2,..., x_n)$ belong to the feasible region $S \subset \Re^n$ [8].

MOEAs find the optimal decision vectors, exploiting the dominance relationship for comparing the candidate solutions, not integrating the individual objective functions as a single function. The dominance relationship can be defined as follows [6].

$$\forall i \in \{1,2,..., m\},\ f_i(\vec{x}_1) \leq f_i(\vec{x}_2) \tag{12}$$

$$\exists i \in \{1,2,..., m\},\ f_i(\vec{x}_1) < f_i(\vec{x}_2) \tag{13}$$

If a decision vector $\vec{x}_1$ is not worse than $\vec{x}_2$ across all the objective functions, $\vec{x}_1$ is said to dominate $\vec{x}_2$. If there is no dominance between the decision vectors, they are said to be non-dominated each other. If a decision vector, $\vec{x}_1$ is not dominated by all other decision vectors in the whole search space, the vector is defined as Pareto-optimal solution [6]. The non-dominated set of the entire search space is the Pareto-optimal set. The Pareto-optimal set in the objective space is called the Pareto-optimal front.

2.5 Implementation of the MOEA

We used non-dominated sorting genetic algorithm (NSGA-II), one of the most popular MOEAs, for the optimization of our d-step parameters [6]. NSGA-II can handle a number of objectives through ranking by non-dominated sorting procedure (Figure 3(a-1)).

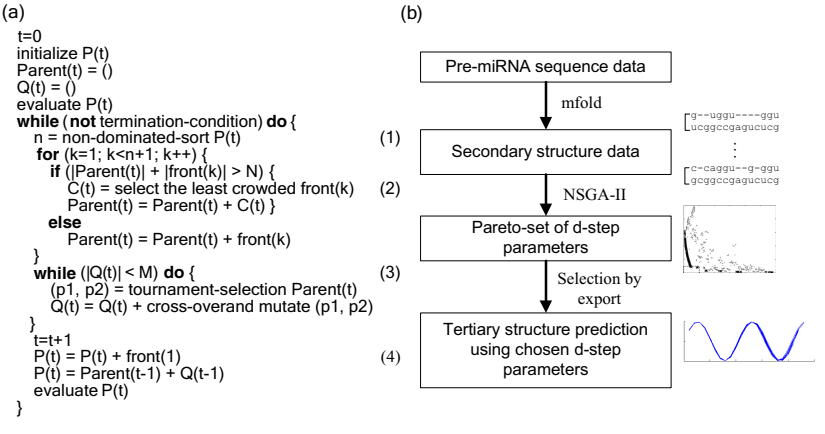

Fig. 3. (a) Pseudo-code of the implemented NSGA-II, (1) creates *n* fronts; (2) produces parents of size *N*; (3) reproduce daughters of size *M*; (4) elitism strategy. (b) Flow chart for tertiary structure prediction using d-step parameters optimized by evolutionary algorithm.

It has shown good convergence and diversity performance on various problem domains. The crowding distance measure in NSGA-II overcomes a drawback of deciding the sharing parameter in the previous NSGA (Figure 3(a-2)). To handle multiple objectives more efficiently, we used the NSGA-II with the elitism strategy (Figure 3(a-4)). The entire NSGA-II procedure is summarized in Figure 3(a).

3 Results

3.1 Case Study: Pre-miRNAs

For the d-step parameterization of pre-miRNA, we used 38 pre-miRNA sequence data where the cleavage site by Drosha/DGCR8 complex is experimentally validated by Northern blotting (http://microrna.sanger.ac.uk/). To extract d-steps of pre-miRNAs, their secondary structures should be known. Here, we applied the mfold package (http://ww.bioinfo.rpi.edu/applications/mfold) for the prediction of secondary structures, and converted the results as pairwise type (Figure 3(b)). We also prepared 20 independent test examples using the same method. The proposed d-step parameter learning procedure is summarized in Figure 3(b).

3.2 Experimental Setting

To find the optimal parameters of dinucleotide steps using NSGA-II, we set several running parameter values as follows. Population size and the number of maximum generation were set to 1000, respectively. As genetic operators, we used the uniform

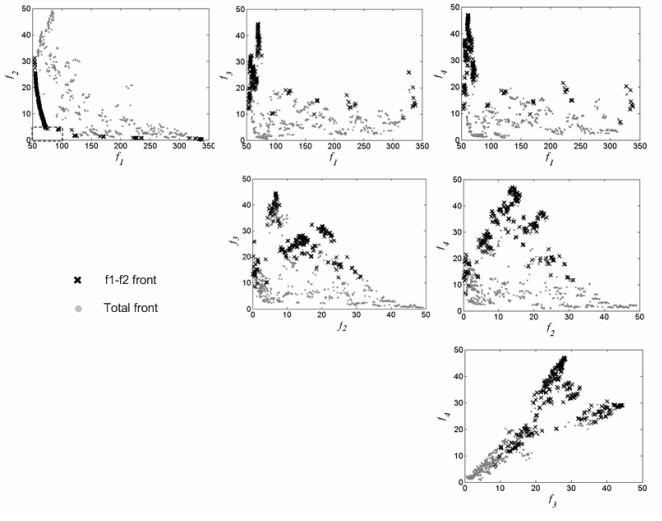

Fig. 4. The relation between the fitness functions. Each graph shows the distribution of total front set of all functions on the pair of functions. The black dots are the front set of function f_1 and function f_2. The gray dots are all front set for all functions.

crossover and 1-point mutation. The rate of crossover was set as 0.9 and the rate of mutation was one over 45, which is the number of total d-step parameters. A chromosome consists of a string with real values of the 45 d-step parameters corresponding to the 3 tertiary structure parameters for each 15 d-steps.

To reduce the search space, we imposed several constraints on the range of each angle (twist, roll, and tilt). For the range values, the angles discovered in DNA double helix were referred to. The twist of d-steps not including deletion is from 20° to 45° and the twist of d-steps including deletion is from 0° to 45°. Roll is from -10° to 10° and tilt is from -5° to 5°.

3.3 The Relation Among the Objectives

Given the running parameter set, we searched for the total front set for all given functions from f_1 to f_4. Figure 4 represents the total front set, a non-dominated set for four functions, which are plotted between pairs of functions. In the results, all pairs of functions excluding the pair of functions f_3 and f_4 show the trade-off relationship. However, the function pair f_3 and f_4 shows a liner relationship, which is reasonable because net roll and tilt are dependent on each other as in Equations 2 and 3.

Among the total front set, we wished to select several sets that are appropriate for our specific purpose. In this study, we focused the twist angle because it is a main contributor of the RNA tertiary structure and thus we first selected the front set determined by functions f_1 and f_2 (black dots in Figure 4). They included 227 parameter sets. Next, we chose eight parameter sets in a specific range where the fitness value of function f_1 is under 100 and the fitness value of function f_2 is under 5 (a small dotted rectangle in the upper left plot in Figure 4).

3.4 Learned Results

The fitness values of the eight selected parameter sets are given in Table 1. Table 1(a) describes the fitness values for the training dataset and Table 1(b) for the test dataset. We concluded that the front sets were optimized well with the given objective functions from the fact that the result on the test dataset is similar to that on the training dataset.

Table 1. The fitness values of each selected front members for training and test datasets

	f_1	f_2	f_3	f_4
1	72.4	4.7	38.3	25.5
2	72.1	4.9	39.7	24.7
3	93.4	4.2	10.2	13.0
4	76.2	4.3	32.1	24.0
5	**93.3**	**4.2**	**10.3**	**13.0**
6	94.2	4.1	10.2	12.8
7	95.0	3.8	10.1	13.4
8	73.0	4.4	35.0	27.2

(a) Fitness values for the training dataset

	f_1	f_2	f_3	f_4
1	82.5	7.1	28.6	21.4
2	83.7	7.7	32.6	18.5
3	82.3	4.1	17.5	14.3
4	90.9	6.0	30.8	20.9
5	**82.5**	**4.0**	**17.5**	**14.4**
6	82.2	4.2	17.2	14.2
7	80.0	4.2	17.5	14.7
8	82.4	6.8	30.4	20.0

(b) Fitness values for the test dataset>

Table 2. The d-step parameter sets of DNA (*) and pre-miRNA. Row is the first nucleotide step and column is the second nucleotide step.

			Purine				Pyrimidine				Deletion
			A	A*	G	G*	U	T*	C	C*	-
Pyrimidine	T	*W*	34.8	40.0							
		R	7.2	2.6							
		T	4.8	0.0							
	C	*W*	32.7	36.9	35.0	31.1					
		R	6.5	1.1	6.1	6.6					
		T	-1.6	0.6	5.0	0.0					
Purine	A	*W*	35.9	35.8	32.9	30.5	34.0	33.4			
		R	6.8	0.5	0.6	2.9	10.0	-0.6			
		T	3.7	-0.4	3.3	-2.0	-0.1	0.0			
	G	*W*	34.9	39.3	34.9	33.4	34.3	35.8	37.6	38.3	
		R	9.6	-0.1	10.0	6.5	9.5	0.4	10.0	-0.7	
		T	5.0	-0.4	4.2	-1.1	4.6	-0.9	2.8	0.0	
Deletion	-	*W*	31.4		35.8		34.6		29.8		34.8
		R	9.4		10.0		6.1		-10.0		9.4
		T	5.0		4.7		4.7		3.9		3.5

3.5 The Representative Parameter Set

Among the eight front parameter sets, we selected a representative parameter set (the fifth parameter set in Table 1) for comparing with the parameter set of DNA, estimated by the crystallography experiment (Table 2). Table 2 demonstrates the 45 d-step parameters of pre-miRNA and the 30 d-step parameters of DNA. It should be noted that there are no deletions in the structure of DNA double-helices and their physical and chemical characteristics are different from those of pre-miRNAs. Nevertheless, several parameters of pre-miRNA have similar values to those of DNA.

3.6 Twist Change of Pre-miRNAs

We drew the twist of pre-miRNAs using the fifth parameter set of Table 2 in the secondary dimension using a sine function (Figure 5). Figure 5(a) displays the twist change for 38 training examples and Figure 5(b) shows the twist change of 20 test examples. The net twists of the training are about 4.1π around 2-turn helices. In both datasets, the variance of net twists increases as the bp-step increases. We guessed that the difference may result from the consecutive deletions after the 10th d-step. For

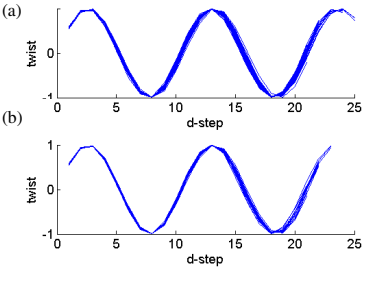

Fig. 5. Twist angles of the training and test datasets calculated using the fifth front parameter set. (a) Twist angles of the training examples (b) Twist angles of the test examples.

example, because '-A(A-)' or '-C(C-)' d-steps have relatively smaller angles than others, their consecutive d-steps will make more difference among twist angles of examples.

4 Discussion

In this study, we suggested an approach to the parameterization of pre-miRNAs using multi-objective evolutionary algorithms (MOEAs). In specific, we optimized the d-step parameters for twist, roll, and tilt angles, needed for predicting the tertiary structure of double helix of pre-miRNA, by introducing four fitness functions. The d-step parameter set of a representative front member had realistic values and shared similar patterns with that of DNA double helix. Also the results on the test dataset were consistent with those on the training dataset with respect to the fitness value and twist change. Combining these, our method has shown to be an effective tool for finding the Pareto-optimal front for this problem.

Multi-objective evolutionary algorithms search for a front, non-dominated optimal set, for given functions. Thus, we should determine the representative results among non-dominated optimal sets in the front. In this study, we mainly focused on the twist angle, a major component of tertiary structure, when choosing a representative front parameter set from the total front sets.

The tertiary structure of a double-stranded RNA is determined by coordinates in the three dimensional space originated form the twist, roll, and tilt angles. In the future work, we will analyze the characteristics of pre-miRNAs on the tertiary structure and investigate the association with the processing mechanism of Drosha/DGCR8 and Dicer RNase III type enzymes. As a matter of fact, binding of proteins or complexes to the DNA or RNA double helices causes conformational changes. The d-step parameters should also be changed by physical or chemical effects of the protein. Hence, dynamic conformational changes can cause difficulties in finding the optimal d-step parameter set. Mitigating the above problem would be another research direction.

Successful improvement of our study may help to understand the processing mechanism of pre-miRNAs as well as to be used as a crucial feature for the prediction of pre-miRNAs and mature miRNAs, which is one of the most challenging problems in miRNA study. These results can also be useful for designing an artificial short hairpin RNA as RNA interference system.

Acknowledgments

This work was supported by the Korea Science and Engineering Foundation(KOSEF) through the National Research Lab. Program funded by the Ministry of Science and Technology (No. M10400000349-06J0000-34910) and the BK21-IT program of the Korea Ministry of Education and Human Resources Development. The ICT at Seoul National University provided research facilities for this study. Kyu-Baek Hwang was supported by the Soongsil University Research Fund. Seong-Bae Park was supported by MIC&IITA through IT Leading R&D Support Project.

References

1. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, (2003). 31(13): p. 3406-15.
2. Olson, W.K., et al., DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A*, (1998). 95(19): p. 11163-8.
3. Strahs, D. and T. Schlick, A-Tract bending: insights into experimental structures by computational models. *J Mol Biol*, (2000). 301(3): p. 643-63.
4. MacKerell, A.D. Empirical force fields: Overview and parameter optimization. in *In 43th Sanibel Symposium*. 2003.
5. Wang, J.a.K., P.A., Automatic parameterization of force field by systematic search and genetic algorithms. *Journal of Computational Chemistry*, (2001). 22: p. 1219-1228.
6. Deb, K.a.G., T. Controlled elitist non-dominated sorting genetic algorithm for better convergence. in *In Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization*. 2001.
7. Zitzler, E., K. Deb, and L. Thiele, Comparison of multiobjective evolutionary algorithms: empirical results. *Evol Comput*, (2000). 8(2): p. 173-95.
8. Mostaghim, S., Hoffman,M., Koenig,P.H., Frauenheim,T., and Teich,J. Molecular Force Field Parameterization using Multi-Objective Evolutionary Algorithms. in *In Proceedings of the Congress on Evolutionary Computation (CEC '04)*. 2004. Portland, U.S.A.
9. Shin, S.-Y., Lee,I.-H., Kim,D. and Zhang,B.-T., Multi-objective evolutionary optimization of DNA sequences for reliable DNA computing. *IEEE Transactions on Evolutionary Computation*, (2005). 9(2): p. 143-158.
10. Lagos-Quintana, M., et al., Identification of novel genes coding for small expressed RNAs. *Science*, (2001). 294(5543): p. 853-8.
11. Kim, V.N., Small RNAs: Classification, Biogenesis, and Function. *Mol Cells*, (2005). 19(1): p. 1-15.
12. Lee, Y., et al., The nuclear RNase III Drosha initiates microRNA processing. *Nature*, (2003). 425(6956): p. 415-9.
13. Kim, V.N., MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol*, (2005). 6(5): p. 376-85.
14. Zeng, Y., R. Yi, and B.R. Cullen, Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *Embo J*, (2005). 24(1): p. 138-48.
15. Schlick, T., *Molecular modeling and simulation*. Interdisciplinary Applied mathematics, ed. S.S. Antman, et al. Vol. 21. 2002, New York: Springer-Verlag.
16. Mathews, D.H., Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics*, (2005). 21(10): p. 2246-53.