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.