

## SECONDARY STRUCTURAL ANALYSIS OF MICRORNA AND THEIR PRECURSORS IN PLANTS

AMIT KUMAR DAS<sup>1</sup>, SAYAK GANGULI\*, SOHINI GUPTA<sup>1</sup> AND ABHIJIT DATTA<sup>2</sup>

<sup>1</sup>DBT-Centre for Bioinformatics, Presidency University, Kolkata

<sup>1</sup>Post Graduate Department of Botany, Barasat Government College, Barasat

<sup>2</sup>Post Graduate Department of Botany, Bethune College, Kolkata

\*Corresponding author. E-mail: sayakbif@yahoo.com

Received: April 19, 2011; Accepted: May 20, 2011

**Abstract**-It has been a general belief over a long period of time that RNA molecules were essential molecules involved only in the process of protein synthesis as carriers of genetic. However recent studies have shown that most of the cells RNAome do not encode proteins. The role and diversity of these numerous non-coding RNAs has not been elucidated till date. The most unique of this plethora of non coding molecules are a class of small RNA molecules of ~ 21 nucleotides in length which has been reported by many workers to be involved in many of the cell process such as controls, defense and development. Short interfering RNAs (siRNA) and Micro RNA (miRNA) represent the two most abundant class of the non coding RNA cascade and both of them are key players in the process of RNA Interference. The current work focuses on the identification and development of secondary structures of available plant microRNA sequences and establishment of a relation between the formation of secondary structures and their free energy values.

**Key words:** siRNA, microRNA, Free energy, Correspondence analysis, hairpin precursors.

### INTRODUCTION

A microRNA (miRNA) is a 21–24 nucleotide (nt) small RNA that is the processed product of a non-coding RNA gene (1,2,3, 5). The mature miRNA is produced from a hairpin structure which lies within the primary transcript (pri-miRNA) and is subsequently processed from this by at least two DCL mediated steps. This processed miRNA is loaded into the RNA INDUCED SILENCING COMPLEX(RISC), in which it performs the silencing process by either binding to the target mRNA and degrading it in a downstream event or by identification and subsequent methylation events.

The first miRNAs in plants was discovered in *Arabidopsis*, and since then more than 700 plant miRNAs have been identified. Generally they have been found to be similar to the animal microRNAs however there are some points of differences.

- Plant pre-miRNAs have been found to possess larger and more variable stem-loop structures.
- Mature plant miRNAs are mostly complementary to their targets.
- Plant miRNAs generally recognize a single target site in the coding region and guide the mRNA to cleavage unlike animal microRNAs which target specific sequences in the 3'UTR.

The specificity of plant miRNA targeting in coding sequences with fewer mismatches suggests that plant miRNAs may act more like siRNAs than do animal miRNAs (4,6).

### MATERIALS AND METHOD

#### DATA MINING AND CLUSTERING

All plant microRNA sequences available were downloaded from miRBASE and the stem loop (precursor) and mature sequences (products) were segregated. The structural analysis were performed using a modified Zuker's algorithm and were then compared to the structures derived using MFOLD. The sequences were then arranged according to their free energies and the total number of stem loop structures were calculated for each sequence. The plant sequences that were used were: *Saccharum officinarum*, *Aegilops tauschii*, *Festuca arundinacea*, *Triticum turgidum*, *Triticum aestivum*, *Hordeum vulgare*, *Sorghum bicolor*, and *Brachypodium distachyon*.

#### CORRESPONDENCE ANALYSISx

Correspondence analysis initially developed for contingency tables, is closely connected with the chi square test for the testing of homogeneity in a contingency table. Various analysis has proved that when there is an association between rows and columns of the table, the value of the underlying chi square statistics is generally high. In CA points are depicted in such a way that the the chi square statistic of the data table is always proportional to the sum of the distances of the points to their centroid (total inertia). The farthest point from the centroid represents the row with the highest contribution to the statistical

value. If we take this into consideration then, CA decomposes the overall chi square statistic. Distances from points are meant to approximate the chi square distance and not the Euclidean distances. When the profiles of the two vectors show similar shape we find that this distance is low, independent of their absolute values.

**RESULTS AND DISCUSSION**

A large number of varied secondary structural conformations were obtained. (Fig 1 and 2) The range of free energies of the mature (- 9.3 to + 1.5) and stem loop sequences (-186.3 to -12.5) were found to conform with the variously reported secondary structure energies of microRNA (FIG 2 A&B). Based on the rigid nucleotide concept Sundaralingam (1969)(8) and reconfirmed by Leontis and Westhof (2003) (6) RNA motifs can be defined as a directed and ordered array of non Watson and Crick base pairs which form clear foldings of the phosphodiester backbone of the interacting strands. In case of microRNAs the identification of secondary structural motifs is important since these molecules have a very short half life and thus the presence of the secondary structures adds to their stability. The folding of RNA molecules have been a debated topic and to date no satisfactory explanation has been put forward to justify the complexities in folding. In this analysis the relationship of free energies and motif formation were tested using a correspondence analysis based on one of the most common secondary structural motif the stem loop (7), and the results indicate that the two variables are closely associated. In both the plots obtained – one for the precursor hairpin loop structures and the other for the mature microRNA sequences we find a clustering around the centroid (FIG 4 A&B) indicating their strong association.

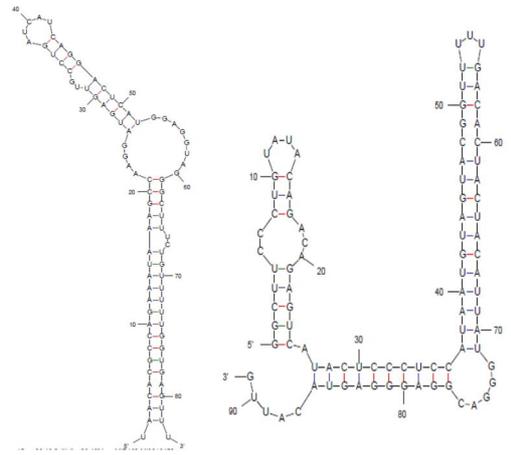


Fig. 1- Representative stem loop structures of the plants under study

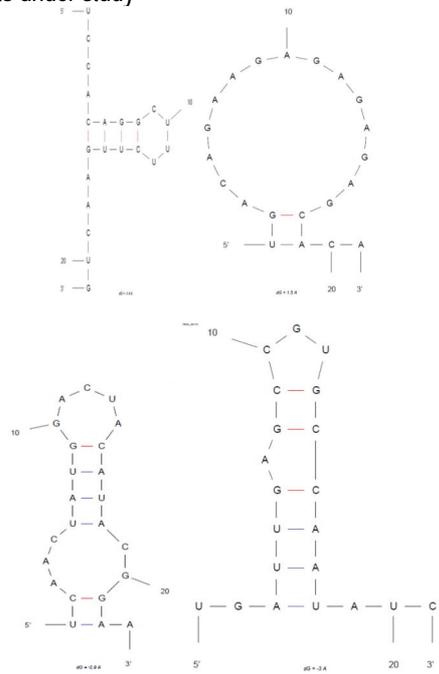
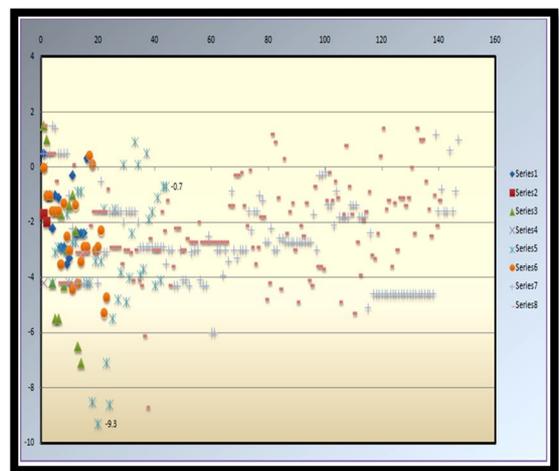
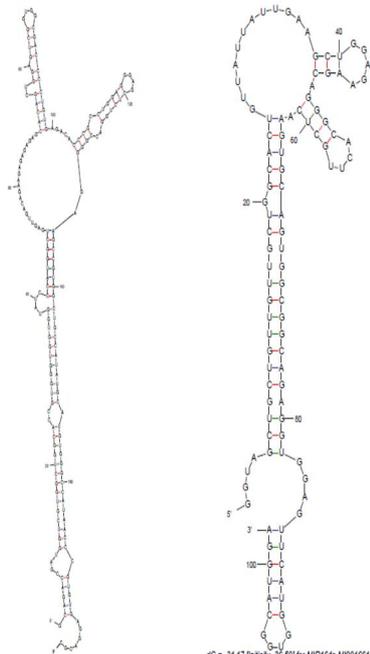


Fig. 2- Representative secondary structures of mature sequences of the plants under study



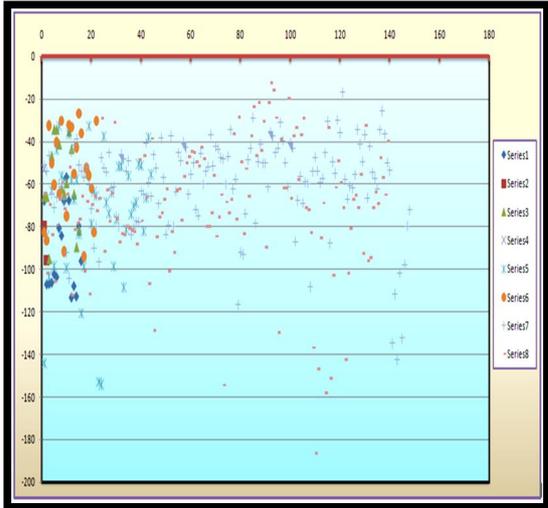


Fig. 3- Graphical scatter plot representations of the energy values of mature and stem loop sequences. [Key: From top to bottom - Series 1=*Saccharum officinarum*, Series 2= *Aegilops taushii*, Series 3= *Festuca arundinacea*, Series 4= *Triticum turgidum*, Series 5= *Triticum aestivum*, Series 6=*Hordeum vulgare*, Series7= *Sorghum bicolor*, Series 8= *Brachypodium distachyon*]

### References

- [1] Fire A., Xu S., Montgomery M.K., Kostas S.A., Driver S.E. and Mello C.C. (1998) *Nature*. 391, 806-811.
- [2] Ganguli S. and Datta A. (2011). *RNA Interference—Interactomics and Therapeutics in RNAi Technology*, edited by V. K. Gupta, CRC Press, 347–356.
- [3] Ganguli S., Dhar P., Basu P., Roy P. and Datta A. (2010) *Intl. J. of Genetics*. 2, 8-17.
- [4] Hammond S.M., Bernstein E., Beach D., and Hannon G.J. (2000) *Nature*. 404, 293-296.
- [5] Kim V.N. (2005) *Molecules and Cells*,19,1-15.
- [6] Leontis B.N. and Westhof E. (2003) *Curr.Opin. Srtuct. Biol.*, 13: 300 – 308.
- [7] Moore P.B. (1999) *Annu Rev Biochem*, 68: 287 – 300.
- [8] Sundaralingam M. (1969) *Biopolymers*, 7:821 – 838.

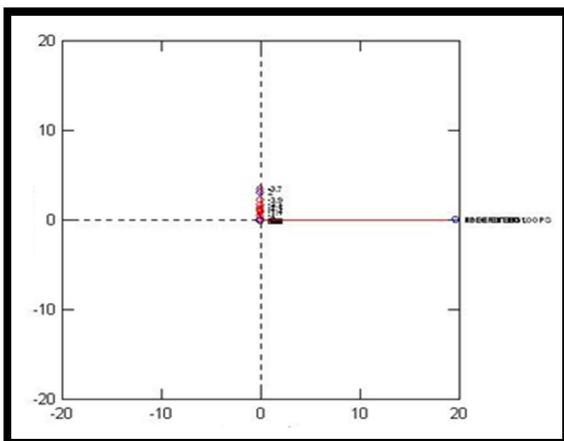
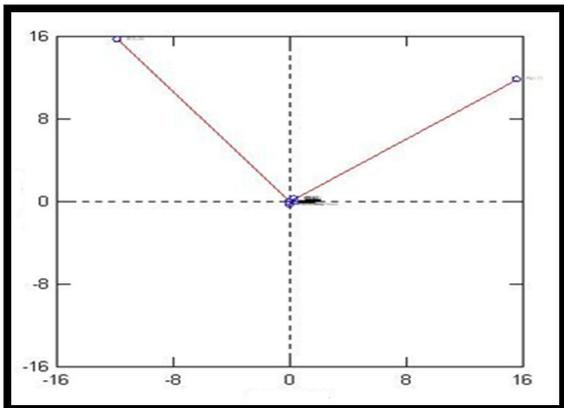


Fig. 4- Correspondence plot of the stem loop and mature sequences of plant microRNAs under study.