I J C B
International Journal of Computational Biology
ISSN: 2229–6700, E-ISSN: 2229–6719, Vol. 2, Issue 2, 2011, pp-35-37
Available online at http://www.bioinfo.in/contents.php?id=96

SECONDARY STRUCTURAL ANALYSES OF MICRORNAS AND PRECURSORS IN *PAN TROGLODYTES*

MOUMITA ADHIKARY, SAYAK GANGULI*, SASTI GOPAL DAS AND ABHIJIT DATTA
DBT Centre for Bioinformatics, Presidency University, 86/1 College Street, Kolkata, India
*Corresponding Author: Email- sayakbif@yahoo.com, Tel- 09830200174

Received: October 08, 2011; Accepted: November 30, 2011

Abstract- *Pan troglodytes* known as the common chimpanzee is an endangered species and finds its mention in the endangered list of animals around the world. Interestingly enough phylogenetic analyses have revealed that this species is an essential sister group to the modern humans and thus is of great evolutionary value. With the increase in data and utilization of micro – RNA based therapeutics comprehensive analyses of micro – RNAs of species having close human lineage is of utmost importance. This work focuses on the secondary structural analyses of micro RNAs and their precursors from *Pan Troglodytes.*
Keywords- *Pan troglodytes*, microRNA, Primary RNA, hairpin precursors, stem-loop structure, and correspondence analysis

Introduction

MicroRNAs (miRNAs) are endogenous, noncoding, small RNAs of approximately 22 nucleotides that function as posttranscriptional regulators. MiRNAs base-pair with the 3'untranslated region complementary sites of the target mRNA, and regulate the stability of the mRNA and translational efficacy. Interest into the field of miRNA field took off after the discovery of the highly conserved let-7 small RNA [9]. It is now known that these tiny molecules serve as important regulators in most organisms for gene expression. More than 400 miRNAs have been identified in humans and are evolutionally conserved Computational predictions have estimated that mammalian miRNAs regulate approximately 30% of all protein-coding genes [7]. MicroRNAs belong to a family of large endogenous RNA. A short single stranded miRNAs are formed in two phases. In nucleus, miRNA transcripts (~60-70nt) are synthesized from Primary miRNA (pri-miRNA) by RNA polymerase II, which is recognized by Drosha-DGCR8 complex (microprocessor) and produces the Precursor miRNA (Pre-miRNA). This Pre-miRNA contains a hairpin shaped structure. Exportin5, a nuclear membrane protein, functions in the export of pre-miRNA into the cytoplasm [3]. Dicer (RNase III) cleaves the pre-miRNA to form short 22nt miRNA with 2nt 3' overhangs. The resultant mature miRNA is recognized by Argonaute protein, Dicer and it finally becomes part of the RNA induced silencing complex (RISC) which results in the cleavage of a complementary mRNA and contributes towards Post-Transcriptional Gene Silencing (PTGS) [4]. But animal miRNA targets are interrupted by gaps, mismatches and 3'UTR of mRNA. Some miRNA are responsible for translational repression [2].

With the advancement in technology RNAi based therapeutics are set to emerge as the next molecular biology revolution and it is for this reason important to identify and characterize the available data in both the secondary and tertiary domains of information.

Materials and Method

Data Mining and Clustering

The work has done on the miRNA sequences of *Pan Troglodytes*, commonly known as Robust Chimpanzee. Chimpanzees are great apes that are most closely related to humans (they share 98% of their DNA with humans). The miRNA sequences were downloaded from miRBase and the stem loop (precursor) and mature (products) sequences were segregated. The structural analysis were performed using a modified Zuker's algorithm and were then compared to the structures derived from M-FOLD. The free energies and the total number of stem loops were calculated for each sequence.

Correspondence Analysis

In correspondence analysis points are depicted in such a way that the chi square statistics 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. 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. [1]
In this analysis the relationship of free energies and motif formation were tested using correspondence analysis based on the presence of one of the most common secondary structural motif- the stem loop [8].

Result and Discussion

The secondary structure of each sequence was obtained and they were based on least free energy (Fig. 1 & 2). In stem loop structures the range of free energy were -61.5 to -9.9 and in mature structures the range were -11.4 to 7.7 (Fig. 3 & 4). Based on the rigid nucleotide concept of [10] and reconfirmed by [11], RNA motifs were a directed and ordered array of non-Watson and Crick base pairs which form clear folding of the phosphodiester backbone of the interacting strands. As the miRNA have a short half-life and the secondary structures give them stability, so the identification of the secondary structural motif is important for miRNA. The debated topic of RNA folding has not been explained satisfactorily till date to justify the folding complexities. The results of correspondence analysis show that the two variables under study are closely associated. In both the plots obtained – one for the precursor hairpin loop structures and the other for the mature microRNA sequences cluster of sequences were observed around the centroid (Fig. 5) indicating their strong association.

Conclusion

A large number of variable structures were obtained indicative of the wide range of structural folding that microRNAs undergo. Future work would focus on the similarities of secondary structures of the conserved groups of microRNAs. The correspondence analysis indicates the relatedness of the number of stem loop structures and the relative entropy of the system.

References

[1] Das K.A, Ganguli S, Gupta S and Datta A. (2011) *IJAS* 3(1) 62 -64.
[2] David T.H. et al (2005) *PNAS*, 102: 16961-16966.
[3] Hutvagner G. (2005) *Federation of European Biochemical Societies*, 579: 5850-5857.
[4] Kim V.N. (2003) *J Korean Med Sci*., 18:309-318.
[5] Lee R.C., Feinbaum R.L., and Ambros V. (1993) *Cell* 75, 843–854.
[6] Leontis N.B. and Westhof E. (2003) *Curr. Opin. Struct. Biol*., 13: 300-308.
[7] Lewis B.P., Burge C.B. and Bartel D.P. (2005) *Cell* 120, 15–20.
[8] Moore P.B. (1999) *Annu Rev Biochem*, 68: 287-300.
[9] Reinhart B.J., Slack F.J., Basson M., Pasquinelli A.E., Bettinger J.C., Rougvie A.E., Horvitz H.R., and Ruvkun G. (2000) *Nature* 403, 901–906.
[10] Sundaralingam M. (1969) *Biopolymers*, 7: 821-838.
[11] Wightman B., Ha I. and Ruvkun G. (1993) *Cell* 75, 855–862.

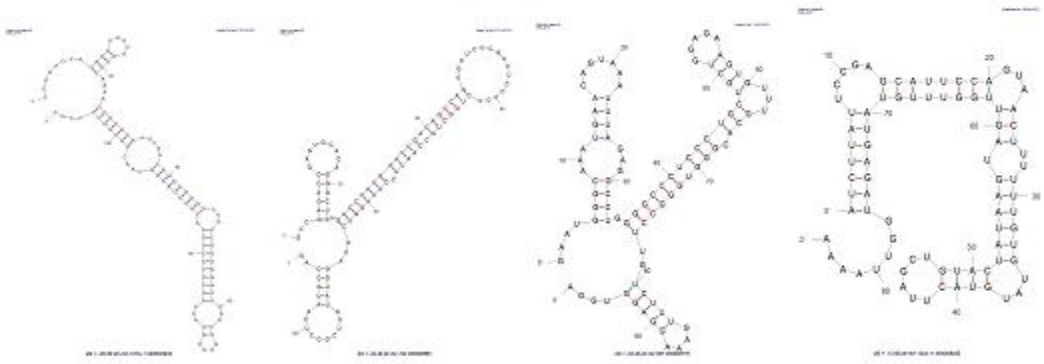

Fig. 1- Representative secondary structures of stem loop sequences of the animal under study

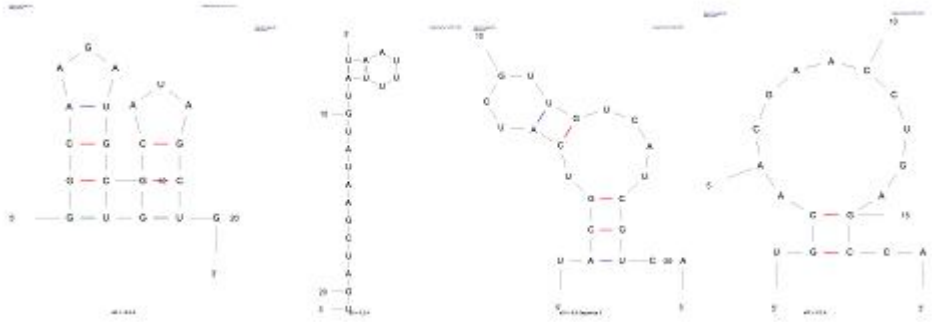

Fig. 2- Representative secondary structures of mature sequences of the animal under study

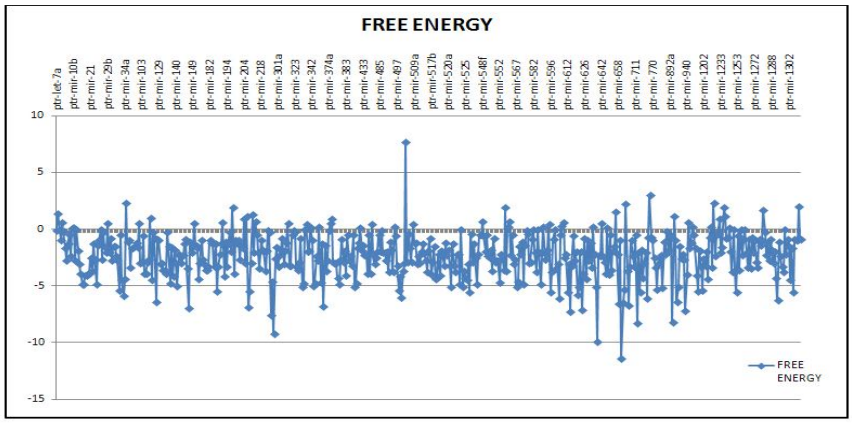

Fig. 3- Graphical representations of the energy values of stem loop sequences

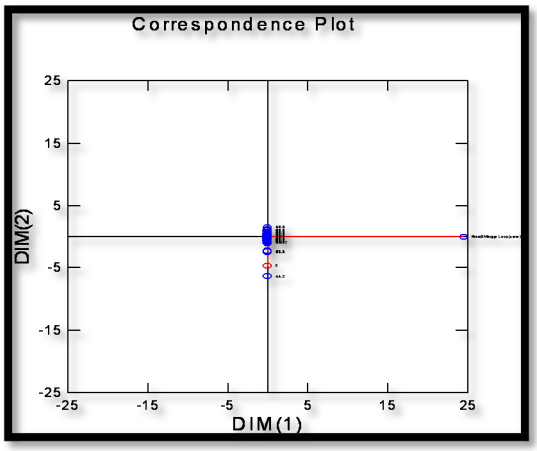

Fig. 4- Graphical representations of the energy values of mature sequences

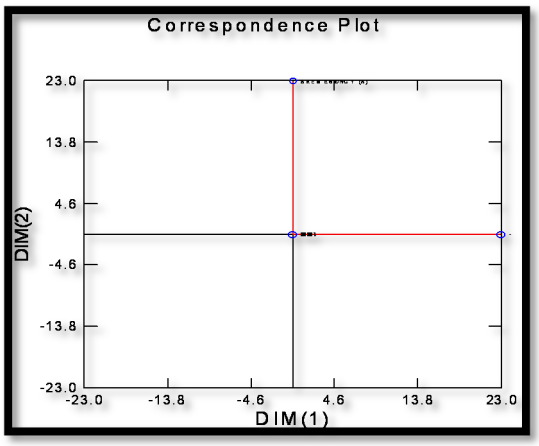

Fig. 5- Correspondence plot of the stem loop (top) and mature sequences (bottom).