International Journal of Computational Bioinformatics and In Silico Modeling
Vol. 3, No. 6 (2014): 549-552
Research Article
Open Access

ISSN: 2320-0634

Computational Analysis of Argonaute Protein Interactions Using Mirror Tree Approach

Pallavi Singh*

Department of Biotechnology, IILM ACADEMY OF HIGHER LEARNING, Greater Noida-201308, India.

*Corresponding author: Pallavi Singh; e-mail: pallavi.singh@iilm.ac.in

Received: 04 November 2014 Accepted: 22 November 2014 Online: 28 November 2014

ABSTRACT

Argonaute proteins have been well characterized for their key role in RNA mediated gene silencing pathways. Association of Argonaute proteins with siRNA to form RISC (RNA induced silencing complex) to catalyze post transcriptional gene silencing by degrading and slicing messenger RNA is established as a significant role of Argonaute protein family. Current work proposes the interaction of Argonaute proteins with DEXD c and Helicase C domain containing protein, RNA helicase LGP2 based on similarity in the phylogenetic distance matrices of the two proteins which is inferred from results obtained by implementation of mirror tree approach. Further verification of this potential interaction has been done by applying different online tools for the prediction of subcellular localization of Argonaute proteins and RNA helicase LGP2. Subcellular localization has shown that like Argonaute proteins, RNA helicase LGP2 is also localized to cytoplasm.

Keywords: Argonaute, Mirror tree, Phylogenetic profiling, RNA Helicase LGP2, Subcellular localization.

INTRODUCTION

Extensive research in last few years over the gene silencing pathways mediated by small RNA molecules has highlighted the role of Argonaute protein family members as key players in RNA interference phenomenon. Small RNAs such as short interfering RNAs (siRNAs), microRNAs (miRNAs) or Piwi-interacting RNAs (piRNAs) are anchored into specific binding pockets and guide Argonaute proteins to target mRNA molecules for silencing or destruction. Various classes of small RNAs and Argonaute proteins are found in all higher eukaryotes and have important functions in processes as diverse as embryonic development, cell differentiation and transposon silencing [1].

Argonaute proteins have conserved domain architecture across various organisms where PAZ domain is present in the middle followed by PIWI domain's occurrence at C-terminal. PAZ domain has been shown to adopt the OB like beta barrel fold confirmation and recognises the 3' end of siRNA, thereby helps in association of Argonaute family proteins with siRNA to form RNA induced gene silencing complex. On the other hand, it has been shown that PIWI domain shares a high degree of topological similarity with RNAse H family of enzymes. Like RNAse H enzymes, PIWI domain also possesses a conserved active site motif comprising of aspartate-aspartate-glutamate residues. This finding has implicitly indicated the role of Argonaute proteins has slicer in RISC which cleaves mRNA and promotes gene silencing by cutting mRNA [2]. It is also known that Ago-1 and Ago-2 proteins interact with RNA helicase MOV10 and the RNA recognition motif (RRM) containing proteins TNR 6B/ KIAA1093 [3].

Based on sequence composition, argonaute protein family is divided into 2 subfamilies in higher eukaryotes- PIWI subfamily and Ago subfamily.8 members of Argonaute protein family are known in Homo sapiens and have been classified into these two subfamilies. *PIWIL1/HIWI, PIWIL2/HILI, PIWIL3,* and *PIWIL4/HIWI2* are categorised into PIWI subfamily and *EIF2C1/hAGO1, EIF2C2/hAGO2, EIF2C3/hAGO3,* and

EIF2C4/hAGO4 are placed into Ago subfamily [4,5]. Despite this classification, domain architecture and presence of PAZ and PIWI domain is conserved across all Argonaute family proteins.

We aimed to study novel protein interactions of Argonaute proteins in higher eukaryotic organisms to decipher other physiological roles and metabolic pathways in which these proteins might participate apart from their known role as slicer in RISC mediated gene silencing phenomenon. Different computational approaches are used to predict protein-protein interaction insilico based on various mechanisms. Phylogenetic profiling, Rosetta stone approach, conservation of gene order, Similarity of Phylogenetic trees, Bayesian network modelling ,Random forest decision tress are some of the well established methods used for insilico prediction of protein-protein interactions [6,7,8,9]. Correlated mutations between the interacting proteins leading to conservation of their interaction interface has been a significant methodology for detecting protein-protein interactions [10]. We have used concept of co-evolution of interacting proteins indicated by similarity in their distance matrices as our basis to predict interacting partners of Argonaute proteins [11,12]. Linear Correlation $r > 0.8$ between the distance matrices of two protein families shows high evolutionary similarity between them indicating their co-evolution and interaction possibility [13,14].

MATERIALS AND METHODS

Finding Domains Occurring Near Argonaute Protein Domains:

Pfam database was searched to see the domain organisation of Argonaute proteins. It was seen that DUF1785 (Pfam ID: PF02170), PAZ (Pfam ID: PF02170) and PIWI (Pfam ID: PF02171) domains occur in sequential manner from N terminus to C-terminus of Argonaute proteins. PAZ and PIWI domains are conserved domains in Argonaute protein family with well defined functions however function of DUF1785 domain is still not known. Domains co-occurring with PAZ and PIWI domain in other eukaryotic proteins were found by checking domain organisation of PAZ and PIWI domains in Pfam database. It was seen that DEAD/DEXD domain, Helicase C and dsRNA domains occur predominantly before the PAZ domain at N-terminal while Ribonuclease III domain is found at the C-terminal of PAZ domain in majority of dicer proteins and Dicer proteins are prominent member of Argonaute protein family. While in case of finding domains occurring near PIWI, it was seen that DUF1785 and PAZ are found at the N-terminal of members of PIWI subfamily belonging to Argonaute protein family. It is known that PAZ domain interacts with PIWI domain and DEAD domain interacts with PAZ domain. Helicase C is conserved domain found at C-terminus of majority of helicase superfamily member though it may not exist as autonomous fold, it is found associated as integral part of helicase family; DEAD, DEXH, DEXDc.

Selection of Probable Interaction Partners of Argonaute Proteins:

Based on the idea that co-occurring domains interact with each other and protein interactions are mediated by these domains, sequence of domains occurring near PAZ and PIWI i.e. sequence of DEAD, DUF1785, Helicase C, dsRNA domains were used to find proteins containing these domains in Homo sapiens in order to find probable interaction partners of argonaute proteins. Moreover, since it is known that PAZ and PIWI domain interact with each other, sequence of these domains were also used to find the other proteins apart from Dicer and Argonaute proteins which contain these domains in order to find putative protein partners of argonaute proteins. Sequence of DUF1785, PAZ and PIWI domains was taken from human Argonaute 1/eiF2C protein (Uniprot accession:) while sequence of DEAD, Helicase C and dsRNA domains was taken from Human endonuclease DICER protein(Uniprot accession:) .Amino acid sequence of these domains was used in BlastP to find proteins in eukaryotic organisms which contain these domains. E-value < .0005 was used as cut off to select eukaryotic proteins which contain these domains. Argonaute protein sequences of higher eukaryotic organisms were taken form Uniprot database. Only those proteins were selected for further analysis by mirror tree approach which co-existed with Argonaute proteins in at least five common higher eukaryotic organisms. Though Pazos,F. and Valencia,A. have suggested criteria of at least 11 common organisms whereby the two proteins should co-exist in order to analyse interactions between two protein families, we restricted our work with cut off of 5 common organisms since we were not able to find at least 11 common organisms having both argonaute protein and other protein obtained from BlastP with specific domain. RNA helicase LGP2, IFIH1/mda5 (Interferon induced with helicase domain1/ Melanoma associated differentiation protein 5), DDX58, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide RIG-I, FANCM, KIAA1567 and HIWI proteins were selected for further analysis.

Multiple Sequence Alignment and Generation of Distance Matrices:

Using Clustal W, multiple sequence alignment of Argonaute proteins and the proteins obtained from BlastP i.e. RNA helicase LGP2, IFIH1/mda5, DDX58, DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide RIG-I, FANCM, KIAA1567 and HIWI proteins was carried using protein sequences of these proteins from organisms which contained both Argonaute proteins as well as these proteins. Multiple sequence alignment result of each protein pair i.e pair of Argonaute protein with LGP2,IFIH1, DDX58, RIG-I,FANCM,KIAA1567 and HIWI protein was analysed one by one after generating a concatenated alignment file of Argonaute with each protein in pir data format and using it as input in Mirrortree software to calculate distance matrices of the Argonaute protein and it's putative protein partner using McLachlan's homology amino acid matrix for finding average homology between amino acid of each

protein family and correlation coefficient between these distance matrices was calculated using Pearson's linear correlation coefficient.

Calculation of Pearson's Correlation Coefficient:
It has been shown in many research works earlier that if Pearson's linear correlation coefficient (r) is equal to or greater than 0.8 between the distance matrices of two protein families, the two proteins share a high degree of evolutionary similarity with each other and they are known to interact with each other based on the concept of co-evolution of interacting proteins.

Prediction of Subcellular Localisation of Rna Helicase Lgp2
Amino acid sequence of RNA Helicase LGP2 from Homo sapiens (Accession No: NP_077024.2) was submitted to three online tools – Psort II, BaCELLO and ESLpred for prediction of subcellular localisation of RNA helicase LGP2.

RESULTS AND DISCUSSION

Distance matrix of LGP2 protein
Command used in mirrortree is:
C:\> Mirrortree.exe LGP2-Agronaute.pir Maxhom_ McLachlan_metric.txt 759 1045 –D

Table 1. RNA Helicase LGP2 Distance matrix

	1	2	3	4	5	6	7	8	9
1		-0.092	-0.061	4.168	0.079	-0.037	3.216	2.801	1.7
2			-0.033	-0.091	-0.069	4.11	-0.025	-0.072	-0.062
3				0.006	-0.038	-0.001	-0.059	-0.116	-0.09
4					0.107	-0.04	3.361	2.908	1.758
5						-0.02	0.018	-0.056	-0.01
6							-0.016	-0.065	-0.05
7								2.604	1.556
8									1.885
9									

Table 2. Distance matrix of Argonaute protein corresponding to RNA Helicase LGP2

	1	2	3	4	5	6	7	8	9
1		-0.217	-0.161	4.315	-0.233	-0.209	4.145	2.85	4.167
2			-0.214	-0.221	-0.177	4.289	-0.201	-0.162	-0.197
3				-0.185	-0.059	-0.267	-0.141	-0.149	-0.137
4					-0.257	-0.199	4.382	2.37	4.415
5						-0.186	-0.224	-0.2	-0.207
6							-0.21	-0.173	-0.216
7								2.208	4.86
8									2.249
9									

Values of Pearson's Linear Correlation Coefficient:

proteins	Correlation ,r
RIG-I	-3
FANCM	-0.35917
KIAA	-0.11124
LGP2	0.90241
HIWI	0.18311
DDX58	-3
IFIH1	0.19015

Based on current work, we propose that Argonaute proteins and RNA helicase LGP2 have co-evolved during course of evolution and they both interact with each other and co-localise in cytoplasm to mediate gene silencing. Negative correlation has been obtained with DEAD/H box containing protein,DDX58 and RIG-I .This indicates that presence of histidine residue along with DEAD box motif , Asp-Glu-Ala-Asp has led to significant evolutionary change between DEAD/H box proteins and Argonaute proteins causing great evolutionary divergence and loss of interaction between these proteins.

CONCLUSION

Observation of the correlation between phylogenetic distance matrices has led to conclusion that argonaute proteins, mainly Ago-1 interacts with DEXD and Helicase C domain containing RNA Helicase LGP2.Moreover it is known that argonaute proteins localise in cytoplasmic P bodies and results obtained in this project indicate that LGP2 is also localized in cytoplasm.

REFERENCES

1. Meister et al. (2008). The Argonaute protein family. Genome Biol. 9:210.
2. Song JJ, Smith SK, Hannon GJ et al. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305: 1434-1437.
3. Meister et al. (2005).Identification of Novel Argonaute associated proteins. Current Biology, 15:2149–2155
4. Takashi Sasaki, Aiko Shiohama, Shinsei Minoshi et al. (2003). Identification of eight members of the Argonaute family in the human genome. Genomics, 82:323-330.
5. Shoemaker BA, Panchenko AR (2007) Deciphering protein–protein interactions. Part II.Computational methods to predict Protein and Domain interaction partners. PLoS Comp Biol 3: e43. doi:10.1371/journal.pcbi.0030043.
6. Ben-Hur A, Noble WS (2005) Kernel methods for predicting protein– protein interactions. Bioinformatics 21: i38–46.
7. Fariselli P, Pazos F, Valencia A, et al. (2002) Prediction of protein–protein interaction sites in heterocomplexes with neural networks. Eur J Biochem 269: 1356–1361.
8. Pazos F, Helmer-Citterich M, Ausiello G et al. (1997). Correlated mutations contain information about protein-protein interaction. J Mol Biol, 271:511-523.

9. Pazos F, Valencia A. (2001).Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein Engineering 14:609-614.
10. Song JJ, Liu J, Tolia NH, et al. (2003).The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol 10:1026-1032.
11. Sato,T. et al. (2005) The inference of protein–protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. Bioinformatics, 21:3482–3489.
12. Tan,S. et al. (2004) ADVICE: Automated Detection and Validation of Interaction by Co-Evolution. Nucleic Acids Res: 32:69–72.
13. Gertz,J. et al. (2003) Inferring protein interactions from phylogenetic distance matrices. Bioinformatics, 19:2039–2045.
14. Goh,C. and Cohen,F.E. (2002). Co-evolutionary analysis reveals insights into protein–protein interactions. J. Mol. Biol., 324:177–192.

*****