

## INTRINSICALLY UNSTRUCTURED REGIONS (IURS) IN ARGONAUTE-1 SEQUENCES ACROSS VARIOUS DOMAINS OF LIFE

PROTIP BASU\*, SAYAK GANGULI, HIRAK JYOTI CHAKRABORTY AND ABHIJIT DATTA

*DBT Centre for Bioinformatics, Presidency University, Kolkata  
Department of Botany, Jhargram Raj College, West Bengal.*

Argonaute proteins have established themselves as key players in the RNAi mechanisms across all domains of life. In plants several types of argonaute proteins are present which are involved in the different small RNA related pathways. Likewise in animals, fungi (Quelling) and Protists argonaute proteins are mostly involved in the microRNA and siRNA related pathways. In some higher eukaryotic taxa they serve as elongation factors as well. This study explores the sequence complexities of Argonaute 1 protein across various domains of life and attempts to identify whether any intrinsically unstructured regions are present in the proteins under study.

During the past decade, small non-coding RNAs have rapidly emerged as important contributors to gene regulation. To carry out their biological functions, these small RNAs require a unique class of proteins called Argonautes. The discovery and our comprehension of this highly conserved protein family is closely linked to the study of RNA based gene silencing mechanisms. With their functional domains, Argonaute proteins can bind small non-coding RNAs and control protein synthesis, affect messenger RNA stability and even participate in the production of a new class of small RNAs, Piwi-interacting RNAs (Hock & Meister, 2008). Plants produce 21-nucleotide siRNAs with the help of DC14, but in its absence, DC12 can substitute, producing 22-nucleotide siRNA. The DC14-produced 21-nucleotide siRNAs typically associate with AGO1 and guide mRNA cleavage. In mammals, AGO1 is occupied by miRNAs, whereas AGO2 associates with siRNAs. miRNAs and siRNAs are chemically indistinguishable and participate in partially overlapping pathways; both are derived from double-stranded RNA (dsRNA) and are then processed into 21-22 nt single stranded molecules by Dicer or a Dicer-like enzyme; later, they are incorporated into the RNA-induced silencing complex (RISC) to guide the cleavage or translational repression of the complementary strand. (Baulcombe 2004, Lindbo and Dougherty 2005) Argonaute proteins are common in many fungi and protists and ubiquitous in plants and animals. The number of Argonaute genes found in these different species varies from one (as in the fission yeast *S. pombe*) to ten (10 in *Arabidopsis*). The eukaryotic Argonaute family can be classified into three major phylogenetic clades based on amino acid sequence similarities (Tolia, and Joshua Tor, 2007). The largest clade is named Argonaute (for clarity, we refer to class this class as Ago) after its founding member AGO1 in *Arabidopsis*. At the cellular level, Ago proteins localize diffusely in the cytoplasm and nucleus and, in some cases, also at distinct foci, which include P-bodies and stress granules (Bohmert *et al.*, 1998, K., Camus *et al.* 2005, Leung *et al.*, 2006).

\*Corresponding author: protipbasu@gmail.com

## INTRINSICALLY UNSTRUCTURED REGIONS (IURS) IN ARGONAUTE-1 SEQUENCES ACROSS VARIOUS DOMAINS OF LIFE

### INTRINSICALLY UNSTRUCTURED REGIONS IN ARGONAUTE-1 SEQUENCES

The second clade, Piwi is named after the *Drosophila* protein PIWI, for P-element-induced wimpy testis, is abundantly expressed in germ line cells and functions in the silencing of germ-line transposons. A diverse assortment of nematode specific Argonaute genes has been grouped together to form a third clade called class 3 or worm Argonautes (WAGOs) (Aravin *et. al* 2006; Yigit *et al.*, 2006)

Hence, by far, AGO1 is the most important Argonaute amongst all Argonaute proteins due to its ubiquitous presence in all domains of life.

### MATERIALS AND METHOD

Sequences were retrieved from the NCBI Genpept database and were subjected to analyses using IUPRED server for detection of the IURs. Once the results were obtained total percentage of IURs were calculated along with the average values for each species (Table 1)

### RESULTS AND DISCUSSION

In all the sequences under study N – terminal Intrinsically Unstructured regions were predominant with a few residues in some of the proteins displaying C – terminal intrinsic unstructuredness. All the plant sequences used in this study showed a large degree of

**Table 1**  
*The set of sequences under study and their intrinsically unstructured regions (IURs).*

SL. NO.	BINOMIAL	KINGDOM	ARGONAUTE LENGTH	AMINO ACIDS IN IURs	% OF IURs	AVERAGE VALUE IURs
1.	<i>Homo sapiens</i>	ANIMALIA	857	32	3.73	0.22
2.	<i>Trypanosoma brucei</i>	PROTISTA	892	104	11.66	0.29
3.	<i>Schizosaccharomyces pombe</i>	FUNGI	834	11	1.32	0.18
4.	<i>Oryza sativa</i>	PLANTAE	1082	263	24.31	0.38
5.	<i>Sorghum bicolor</i>	PLANTAE	1109	300	27.05	0.42
6.	<i>Zea mays</i>	PLANTAE	1078	265	24.58	0.38
7.	<i>Solanum lycopersicum</i>	PLANTAE	1054	221	20.97	0.39
8.	<i>Solanum tuberosum</i>	PLANTAE	1054	220	20.87	0.39
9.	<i>Manihot esculenta</i>	PLANTAE	993	166	16.72	0.32
10.	<i>Gossypium raimondii</i>	PLANTAE	1080	257	23.8	0.38
11.	<i>Arabidopsis thaliana</i>	PLANTAE	1048	226	21.56	0.37
12.	<i>Brassica rapa</i>	PLANTAE	1079	257	23.82	0.39
13.	<i>Glycine max</i>	PLANTAE	1058	228	21.55	0.37
14.	<i>Phaseolus vulgaris</i>	PLANTAE	1063	226	21.26	0.37

## INTRINSICALLY UNSTRUCTURED REGIONS IN ARGONAUTE-1 SEQUENCES

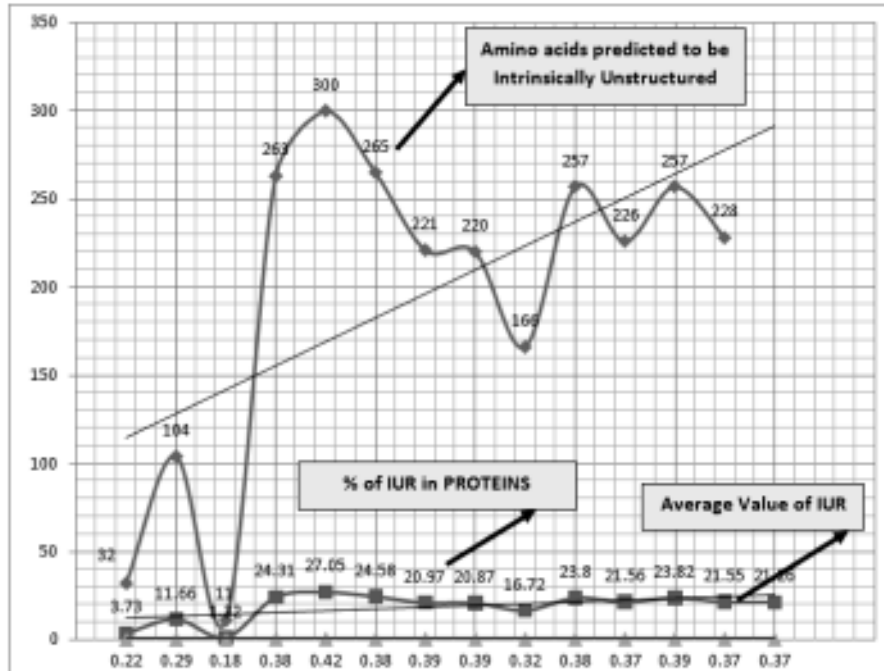


Fig 1

Graphical representation showing the uniformity of distribution of average intrinsically unstructured regions in Argonaute – 1 sequences across the species under study.

unstructured regions in their N – terminal regions. Most of such regions (Table 1; Fig 1) had a high Glycine bias, thus rendering them flexible to fold in any secondary structure deemed necessary for the protein to function.

### CONCLUSION

The 3D structures of argonaute 1 sequences used in this study should provide us with palpable insights regarding the actual significance of the intrinsically unstructured regions in future.

### ACKNOWLEDGEMENT:

The authors acknowledge the financial contributions of the UGC – JRF fellowship scheme and the DBT – BIF – BTBI scheme.

REFERENCES

- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein M J, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo J J, Ju, J, Sheridan, R, Sander, C, Zavolan, M, and Tuschl, T (2006): *Nature* **442**, 203–207.
- Baulcombe D (2004): RNAi silencing in plants. *Nature*, **431**, 356-363.
- Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, and Benning C. (1998): *EMBO J.* **17**, 170–180.
- Hock J and Meister G (2004): The Argonaute protein family. *Genome Bio.* **9**, 210.
- Hutvagner G, Simard MJ (2009): Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**, 22-32.
- Leung A K, Calabrese J M, and Sharp P A (2006): *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18125–18130
- Lindbo J, Dougherty W (2005): Plant Pathology and RNAi: A Brief History. *Annu Rev Phytopathol* **43**, 191-204.
- Liu J, Valencia-Sanchez M A, Hannon G J, and Parker R. (2005) *Nat. Cell Biol.* **7**, 719–723,
- Tolia N H, and Joshua-Tor L (2007): *Nat. Chem. Biol.* **3**, 36–43
- Yigit E, Batista P J, Bei Y, Pang K M, Chen C C, Tolia N H, Joshua-Tor L, Mitani S, Simard M J, and Mello C C (2006): *Cell* **127**, 747–757.