Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Biomedicine

journal homepage:www.elsevier.com/locate/apjtb

Document heading doi:10.1016/S2221-1691(11)60042-7

Modeling and analysis of *Schistosoma* Argonaute protein molecular spatial conformation

Jianhua Zhang[1], Zhigang Shang[1], Xiaohui Zhang[2], Yuntao Zhang[3*]

[1]*The Faculty of Biomedical Engineering of Zhengzhou University, Zhengzhou 450001, Henan Province, China*
[2]*College of Medicine, University of South Florida, FL 33647, USA*
[3]*Department of Urology Surgery, Xijing Hospital, the Fourth Military Medical University, Xi`an, Shan´xi, China*

ARTICLE INFO

Article history:
Received 20 March 2011
Received in revised form 2 April 2011
Accepted 16 April 2011
Available online 30 April 2011

Keywords:
Argonaute protein
Distribution of secondary structure
Phyre
Domain
Schistosoma
Spatial conformation
Amino acid sequence

ABSTRACT

Objective: To analyze the amino acid sequence composition, secondary structure, the spatial conformation of its domain and other characteristics of Argonaute protein. **Methods:** Bioinformatics tools and the internet server were used. Firstly, the amino acid sequence composition features of the Argonaute protein were analyzed, and the phylogenetic tree was constructed. Secondly, Argonaute protein's distribution of secondary structure and its physicochemical properties were predicted. Lastly, the protein functional expression form of the domain group was established through the Phyre-based analysis on the spatial conformation of Argonaute protein domains. **Results:** 593 amino acids were encoded by Argonaute protein, the phylogenetic tree was constructed, and Argonaute protein's distribution of secondary structure and its physicochemical properties were obtained through analysis. In addition, the functional expression form which comprised the N-terminal PAZ domain and C-terminal Piwi domain for the Argonaute protein was obtained with Phyre. **Conclusions:** The information relationship between the structure and function of the Argonaute protein can be initially established with bioinformatics tools and the internet server, and this provides the theoretical basis for further clarifying the function of *Schistosoma* Argonaute protein.

1. Introduction

The life history of the *Schistosoma* is complex with alternating sexual and asexual reproduction as well as converting intermediate host and the definitive host. The accurate expression and the precise regulation and control of the *Schistosoma* genes are the premise of completing the complex life history. Over the past 10 years, studies have shown that non-coding RNA plays an important role in regulating and controlling the growth and development of the living body, and Argonaute protein is the key molecule involved in this process[1-3].

At present, there have already been reports on Argonaute protein in the studies on *Schistosoma japonicum*. The *in vivo Schistosoma* gene knock-down can be successfully realized with *in vitro* synthesized small interference RNA (siRNA) molecules, and the *Schistosoma* expresses a series of microRNA (miRNA) molecules in different developmental stages, which suggests that Argonaute protein plays an important role in the growth and development of *Schistosoma*[4-6]. In view of this, this study has laid the initial foundation for analyzing the *Schistosoma* Argonaute protein's structural characteristics with bioinformatics tools and the internet server, and taking this as a starting point, for developing new anti-schistosomal agent target and vaccine molecules.

2. Materials and methods

2.1. Polypeptide sequence of Argonaute protein

Through looking up the *Schistosoma japonicum* database (http://www.ncbi.nlm.nih.gov/BLAST) in GenBank, one *Schistosoma japonicum* EST fragment (GenBank:AAW26476.1) of 593aa was obtained, and its FAST format was:

>gi|56756607|gb|AAW26476.1| SJCHGC01111 protein [*Schistosoma japonicum*]
MYEKYGDNMARCSTQMAHDLRRIRVETEKFYKSDNGNAYSRRFTVHG
ISSVPANQLMIEELKQSVAAYFDQHHHIKLKYPELPCVKVDQKREVYM
PMELLNILPYQAPNASKAEVASEVVRCAAVRPQERFQELQTFANNMLK
SHPLINQFGLTVQPRPVVVNARVIHPPSAAFGRSHVVPLKAGSWTSPGF
HDPAGRGVELLWAILCVPPDRRSQGHIQKVMHELPRAADRVGMRLSS

*Corresponding author: Yuntao Zhang, Department of Urology Surgery, Xijing Hospital, the Fourth Military Medical University, Xi`an, Shan´xi, China 710032.
Email: peter_hua@sohu.com
Foundation Project: Supported by the the natural sciences plan projects of education department of Henan province (No: 2010A310018)

RPSVSQCPTGELNRRFDEFSRQGCSFLLLILYDEYSYPTIKRLSDLQMGIR TQCVRGRTLDKPNVFPNLLLKLNGKLGGVNWQIPDLIKNGNELIMVFGA DVTHPAPTQNQQIRKSVAAVIGSVSPDLMRYGVVIRQQATTEKGNKTAR EIIDDMRLIVKELLQLYLRNTNGRFPNRMIFYRDGVSEGQFENVLVEELA AIQRACADVRPGEEPAITYIVVQKRHHIRFKPSDPRARNVEPGTIVDTEI THPREFDFYLCSQDGIQGTSKPAHYHVLYDDSNWTSDALQMFTYYLCY AYMRCSRSVSYPAPTYYSHLAAFRARDWLSGMDQPSALLDCGRFKVHM SQVDGMFYL

2.2. Primary sequence and its phylogentic analysis

NCBI (http://www.ncbi.nlm.nih.gov/BLAST) Blastp was used to analyze and predict the homology. DNAStar software was used to analyze the number of amino acid residues, composition and the protein's relative molecular mass and other parameters. According to the Ago family proteins of *Drosophila*, mouse, human and *Schistosoma mansoni* logging on the protein database, including Dm.Ago1 (NP_725342.1), Dm.Ago2 (ABB54719.1), Dm.Ago3 (ABO26294), Ce.ALG1 (NP_510322.2), Ce.ALG2 (NP_493837.1), Hs.AGO1 (Q9UL18), Hs.AGO2 (Q9UKV8), Hs.AGO3 (Q9H9G7.2), Hs.AGO4 (Q9HCK5), Mm.AGO2 (Q8CJG0.2), Mm.AGO1 (Q8CJG1), Mm.AGO3 (Q8CJF9), Mm.AGO4 (Q8CJF8), Sm.Ago1 (Smp_140010), Sm.Ago2 (Smp_179320), Ago3 (Smp_102690.2) and Sm.Ago4 (Smp_102690.3), the ClustalX software was used to make multiple alignment of the Argonaute protein amino acid sequence of different species, and MEGA4 software was further used to construct the phylogenetic tree.

2.3. Secondary structure prediction and physicochemical property analysis

The primary structure of Argonaute protein was analyzed, and the Argonaute protein consisted of 593 amino acids, containing 20 strongly acidic amino acids, 23 strongly cationic amino acids, 97 hydrophobic amino acids and 83 polar amino acids, and the relative molecular mass was 22609.52. The Phyre-based molecular structure modeling was analyzed. Based on the Chou-Fasman and Gamier-Robson algorithms, the module provided by the Predictprotein network resource was used to predict the secondary structure and its distribution of Argonaute protein. DNAStar software and Kyte-Doolittle program were applied to predict its hydrophilicity, Emini was used to predict the surface accessibility, and Karplus-Schulz program was used to predict its flexibility.

In the absence of homology information, the threading was selected to predict the molecular folding structure of Argonaute proteins, and the basic principle was to "thread" the Argonaute protein's amino acid sequence information into the basic framework of the known protein to predict the spatial conformation of Argonaute protein domain by calculating the probability of various folding. The Phyre operation used Profile-Profile alignment and the secondary-structure three-dimensional automatic modeling server (maintained by Imperial College London), and it firstly used the standard PSI-Blast to make multiple sequence alignment, then conducted structure alignment of a series of members in the family to generate its sequence spectrum, and subsequently, the threading was used to make alignment of 1D-3D sequence spectrum between Argonaute protein molecules and the template.

3. Results

3.1. Argonaute protein's amino acid sequence

It was shown through multiple alignment and analysis of sequences that the encoding amino acid sequence of *Schistosoma japonicum* Argonaute protein (GenBank:AAW26476.1) had a high homology of 90% with the XP_002581078.1, of 83% with the AAW25407.1, and of 38% with the XP_001975706.1. The analysis of the phylogenetic tree indicated that the *Schistosoma japonicum* Argonaute obtained in this study was closer to *Drosophila* (Figure 1).

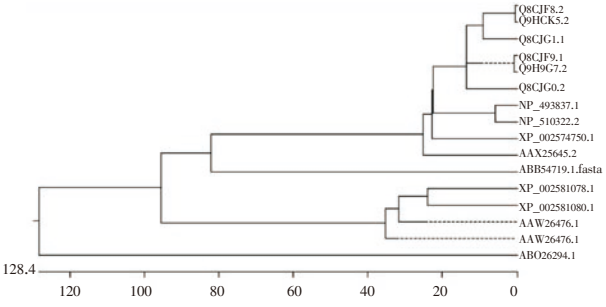

Figure 1. The phylogenetic tree of the *Schistosoma japonicum* Argonaute.

3.2. Argonaute protein's secondary structure

Predictprotein (http://www.predictprotein.org/) was used to predict the secondary structure. The results showed that Argonaute protein belonged to a hybrid protein, in which the helix accounting for 32.83%, the folding accounting for 19.70% and the loop accounting for 47.47%, and the predicted secondary structure and its distribution of Argonaute protein was shown in Figure 2.

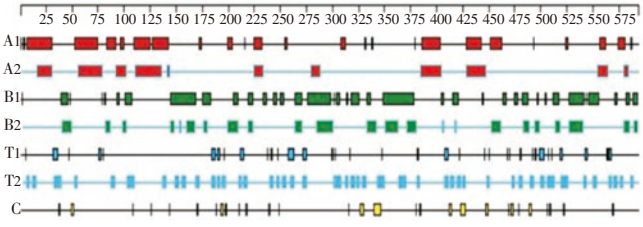

Figure 2. The predicted distribution of secondary structure of Argonaute protein.

A1: Alpha, Regions-Gamier Robson; A2: Alpha, Regions-Chou-Fasman; B1: Beta, Regions- Gamier-Robson; B2: Beta, Regions-Chou-Robson; T1: Turn, Regions-Gamier-Robson; T2: Turn, Regions-Chou-Robson; C: Coil, Regions-Chou-Robson.

3.3 Argonaute protein's physicochemical properties

The distribution of Argonaute protein's hydrophilic regions was uneven, and the distribution in the central part was more intensive, being the high hydrophilic region

(Figure3). Argonaute protein's accessibility analysis curve was similar to the hydrophilicity program analysis curve (Figure 3). The regions of high accessibility were on the surface of Argonaute protein molecules, and the regions of low accessibility were buried in the intramolecular regions. These regions might have certain flexibility, and it was much likely to form the epitope, which was easy to be chimeric with the antibody (Figure 3).

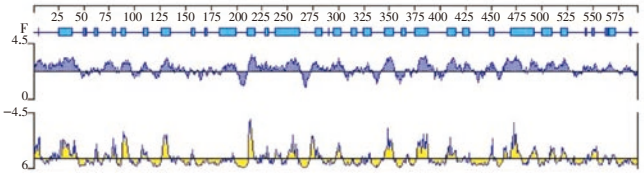

Figure 3. Prediction results of Argonaute protein's physicochemical properties.
● Surface Probability Plot-Emini; ● Hydrophilicity Plot-Kyte-Doolittle; ● Flexible Regions-Karplus-Schulz.

3.4. Argonaute protein molecular spatial conformation

By using Profile-Profile alignment and the secondary-structure three-dimensional automatic modeling server, Phyre "threaded" the Argonaute protein's amino acid sequence information into the basic framework of the known protein, and the functional expression form comprised the N-terminal PAZ domain and C-terminal Piwi domain for the Argonaute protein domain's spatial conformation was obtained by calculating the probability of various folding (Figure 4).

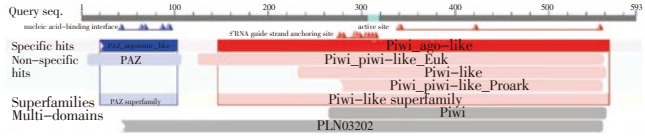

Figure 4. The functional expression form consisted of the Argonaute protein.

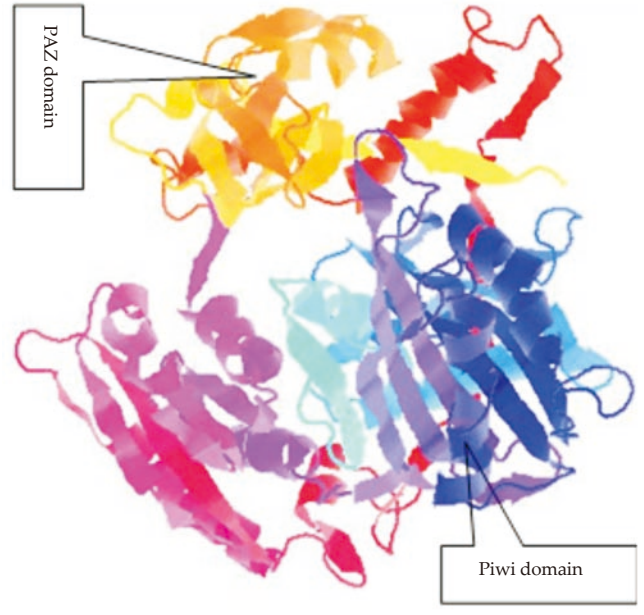

Figure 5. Argonaute protein molecular spatial conformation prediction.

The PAZ domain provides a unique mode for the recognition of the two 3′-terminal nucleotides in single-stranded nucleic acids and buries the 3′ OH group. It might recognize characteristic 3′ overhangs in siRNAs within RISC (RNA-induced silencing) and other complexes (Figure 5). The Piwi domain is the C-terminal portion of Argonaute and consists of two subdomains, one of which provides the 5′ anchoring of the guide RNA and the other, the catalytic site for slicing (Figure 5).

4. Discussion

Studies in recent years have shown that the non-coding small RNA, such as siRNA and miRNA, and Piwi interacting RNA (piRNA) play an important regulating and controlling role in the growth and development of organisms[7-9], and Argonaute plays a major role in the regulation and control process through selectively binding with small RNA. Argonaute protein was firstly found in plants, then such protein molecules were found in yeast, nematodes, mouse and human and many other species subsequently, and they usually contain two functional domains of PAZ (PIWI-Argonaute-Zwille) and Piwi. Therefore, Argonaute protein is divided into two subfamilies of PAZ and Piwi.

The life history of the *Schistosoma* is complex with alternating sexual and asexual reproduction as well as converting intermediate host and the definitive host, and the accurate expression and the precise regulation and control of the *Schistosoma* genes is the premise of completing the complex life history. Over the past 10 years, studies have shown that non-coding RNA plays an important role in regulating and controlling the growth and development of the living body, and Argonaute protein is the key molecule involved in this process. Cheng *et al* and Yang *et al*[10,11]. looked up the *Schistosoma* expressed sequence tag database according to the mouse and human Argonaute protein sequences, and used 5′RACE technique to amplify one expressed sequence tag. Bioinformatics analysis showed that the full-length of cDNA sequence of a coding *Schistosoma japonicum* Argonaute protein was obtained, and the conserved functional domain was analyzed to contain PAZ and Piwi domains, which was the typical domain feature of Argonaute protein. The researchers obtained the full-length cDNA sequence of a coding *Schistosoma japonicum* Argonaute protein with 5′RACE technique combined with bioinformatics study, and conducted the procaryon expression of its encoded protein[11]. In this study, the bioinformatics was used to analyze the domain spatial conformation of *Schistosoma* Argonaute protein, and this laid the initial foundation for further study on the function of Argonaute protein in the growth and development of *Schistosoma*.

According to relevant literature[12-15], the *Schistosoma* Argonaute protein is a highly conserved family member, and it is a multi-domain protein which contains the N-terminal domain, PAZ domain, MID domain and Piwi domain. Through the analysis in this study, the functional expression form comprised the N-terminal PAZ domain and C-terminal Piwi domain for the Argonaute protein domain's spatial

conformation was obtained. The PAZ domain provides a unique mode for the recognition of the two 3′-terminal nucleotides in single-stranded nucleic acids and buries the 3′ OH group, and that it might recognize characteristic 3′ overhangs in siRNAs within RISC (RNA-induced silencing) and other complexes. The PIWI domain is the C-terminal portion of Argonaute and consists of two subdomains, one of which provides the 5′ anchoring of the guide RNA and the other, the catalytic site for slicing. The two landmark domains-the N-terminal PAZ domain and C-terminal Piwi-in Argonaute protein family were obtained with the structural homology-based threading folding method.

In recent years, with the rapid development of bioinformatics, biologists have designed a lot of molecular biological software for protein analysis[16-19]. The analysis on the transcription of Argonaute protein in different development stages made by Yang *et al*[11]. It showed that by combining with the non-coding RNA, this protein could be coordinately involved in the regulation and control of the growth and development of *Schistosoma japonicum*; it was also shown through multiple sequence comparison and analysis that *Schistosoma japonicum* Argonaute protein had relatively high homology with *Schistosoma mansoni* Argonaute protein[20]. In this study, bioinformatics tools and the internet server were used to predict the *Schistosoma* Argonaute protein′s distribution of secondary structure and its domain spatial conformation firstly, and the results showed that the *Schistosoma* Argonaute protein contained α-helix and β-folding, belonging to a hybrid protein structure; the conserved functional domain of coding *Schistosoma japonicum* Argonaute protein was obtained, and it was analyzed to contain PAZ and Piwi domains, which was the typical domain feature of Argonaute protein. This study has laid the initial foundation for further study on the function of Argonaute protein in the growth and development of *Schistosoma*, and has also provided the theoretical basis for further clarifying the function of *Schistosoma* Argonaute protein.

Conflict of interest statement

We declare that we have no conflict of interest.

References

[1] Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, et al. Recognition of 2′-O-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. *Structure* 2011; **19**(2): 172-180.

[2] Moskwa P, Buffa FM, Pan Y, Panchakshari R, Gottipati P, Muschel RJ, et al. miR-182-mediated downregulation of BRCA1 impacts DNA repair and sensitivity to PARP inhibitors. *Mol Cell* 2011; **41**(2): 210-220.

[3] Mallory A, Vaucheret H. Form, function, and regulation of Argonaute proteins. *Plant Cell* 2010; **22**(12): 3879-3889.

[4] Zhao S, Wang Y, Wei H. Trichosanthin induced Th2 polarization status. *Xibao Yu Fenzi Mianyixue Zazhi* 2006; **3**(4): 297-301.

[5] Zhou H, Jiao Z, Pan J, Hong J, Tao J, Li N, et al. Immune suppression via IL-4/IL-10-secreting T cells: a nontoxic property of anti-HIV agent trichosanthin. *Clin Immunol* 2007; **122**(3): 312-322.

[6] An Q, Wei S, Mu S, Zhang X, Lei Y, Zhang W, et al. Mapping the antigenic determinants and reducing the immunogenicity of trichosanthin by site2directed mutagenesis. *J Biomed Sci* 2006; **13**(5): 637-643.

[7] Chen J, Yang Y, Guo S, Peng J, Liu Z, Li J, et al. Molecular cloning and expression profiles of Argonaute proteins in *Schistosoma japonicum*. *Parasitol Res* 2010; **107**(4): 889-899.

[8] Hao L, Cai P, Jiang N, Wang H, Chen Q. Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics* 2010; **11**: 55.

[9] Krautz-Peterson G, Bhardwaj R, Faghiri Z, Tararam CA, Skelly PJ. RNA interference in schistosomes: machinery and methodology. *Parasitology* 2010; **137**(3): 485-495.

[10] Chen J, Guo SX, Yang YP, Liu JM, Lin JJ, Li JK, et al. Cloning and preliminary analysis of full-length cDNA encoding Argonaute 3 from Schistosoma japonicum. *Zhongguo Xuexichongbing Fangzhi Zazhi* 2010; **4**: 310-314.

[11] Yang YP, Guo SX, Chen J, Lin JJ, Liu ZP, Cheng GF. Cloning, expressing and identifying of a full-length cDNA encoding Argonaute from Schistosoma japonicum. *Zhongguo Renshou Gonghuanbing Zazhi* 2010; **9**: 830-834.

[12] Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* 2010; **11**(7): 522-527.

[13] Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. Structure of the guide-strand-containing argonaute silencing complex. *Nature* 2008; **456**(7219): 209-213.

[14] Mallory AC, Hinze A, Tucker MR, Bouché N, Gasciolli V, Elmayan T, et al. Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. *PLoS Genet* 2009; **5**(9): e1000646.

[15] Wang Y, Li Y, Ma Z, Yang W, Ai C. Mechanism of microRNA-target interaction: molecular dynamics simulations and thermodynamics analysis. *PLoS Comput Biol* 2010; **6**(7): e1000866.

[16] Xu YF, Liu WL, Dong JQ, Liu WS, Feng QS, Chen LZ, et al. Sequencing of DC-SIGN promoter indicates an association between promoter variation and risk of nasopharyngeal carcinoma in cantonese. *BMC Med Genet* 2010; **11**: 161.

[17] Balasubramanian C, Ojha RP, Maiti S, Desideri A. Sampling the structure of the noncanonical lin-4:lin-14 microRNA: mRNA complex by molecular dynamics simulations. *J Phys Chem B* 2010; **114**(49): 16443-16449.

[18] Liu X, Liu S, Li H, Han Z, Shao Y, Kong X. Unique sequence characteristics of genes in the leftmost region of unique long region in duck enteritis virus. *Intervirology* 2009; **52**(5): 291-300.

[19] Gupta AR, Patra RC, Das DK, Gupta PK, Swarup D, Saini M. Sequence characterization and polymerase chain reaction-restriction fragment length polymorphism of the mitochondrial DNA 12S rRNA gene provides a method for species identification of Indian deer. *Mitochondrial DNA* 2008; **19**(4): 394-400.

[20] Gomes MS, Cabral FJ, Jannotti-Passos LK, Carvalho O, Rodrigues V, Baba EH, et al. Preliminary analysis of miRNA pathway in *Schistosoma mansoni*. *Parasitol Int* 2009; **58**(1): 61-68.