

THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Characterisation and Functional Analysis of A0A022RDQ0 Protein

Anakaa Tavershima Moses

Assistant Lecturer, Department of Biological Sciences, Federal University Dutsin-Ma, Katsina State, Nigeria

Bem Alexander Adi

Professor, Department of Biological Sciences, Federal University Dutsin-Ma, Katsina State, Nigeria

Musa Danladi Daniel

Lecturer, Department of Biological Sciences, Federal University Dutsin-Ma, Katsina State, Nigeria

Abstract:

This work seeks to characterise and carry out functional analysis of A0A022RDQ0 protein. Basic Local Alignment Search Tool was used to consider specific similarity hits. Uniprot database was used to check predicted functional similarities and subcellular localization of model proteins. Clustal Omega database was used to obtain phylogenetic tree of all proteins to determine evolutionary relationships. Domain/superfamily/multi-domain hits of model proteins were analysed using NCBI's conserved domain database. Results report that A0A0B2RQU6, G7L0J5, W9SFC3 and A0A0B0P4Q5 proteins respectively show 85.5%, 84.0%, 84.09% and 82.4% similarities to A0A022RDQ0. The four model proteins are ribonucleoproteins that localise in the ribonucleoprotein complex. Phylogenetic tree report that all proteins have similar evolutionary origin. Domain search indicate that all proteins belong to WD40. High similarity in sequence of A0A022RDQ0 to four model proteins as well as similar function, domain, subcellular localization and evolutionary origin of the various proteins suggests that A0A022RDQ0 functions as a ribonucleoprotein.

Keywords: Characterisation, Functional analysis, A0A022RDQ0 protein

1. Introduction

A0A022RDQ0 is an uncharacterized protein of the yellow monkey flower, *Mimulus guttatus*. This protein consists of 339 amino acids and comprises of WD40 domain that is common in nearly all eukaryotic proteins (Finn et al., 2016). WD40 performs many functions that include cytoskeleton assembly, apoptosis, pre-mRNA assembly, transcription and signal transduction via regulatory modules/adapters. The major function of WD40 domain is believed to be multi-protein assemblies where tandem repeats function as rigid scaffold in protein interactions. Such assemblies include E3 Ubiquitin ligase and G proteins. This domain consists of 11-24 dipeptide residues called the GH and another with a length of 40 residues at the C terminus that accounts for its name WD40. The 40th amino acid residue usually ends with tryptophan-aspartic acid which is a conserved core between the GH and WD that make stable and reversible binding of protein possible. This is attributed to the formation of a circular structure that is propeller-like and comprises of many blades that are made up of four stranded antiparallel beta sheets (Finn et al., 2016; Hellsten et al., 2013; Marchler-Bauer et al., 2015, 2011, 2009, 2004).

The function of A0A022RDQ0 protein is unknown. This work seeks to characterise and carry out functional analysis of A0A022RDQ0 protein. The specific objectives of this work are; to identify proteins with specific similarity hits, to explore predicted functional similarities and subcellular localization of the model proteins as an insight to that of A0A022RDQ0, to check phylogenetic tree of all proteins to determine evolutionary relationship and to perform domain/superfamily/multi-domain hits to confirm functional similarity of the five proteins.

2. Materials and Methods

Basic Local Alignment Search Tool was used to consider specific similarity hits. Uniprot database was used to check predicted functional similarities and subcellular localization of model proteins. Multiple sequence alignment of A0A022RDQ0 with corresponding proteins identified was done using Clustal Omega database to check phylogenetic tree of all proteins to determine evolutionary relationship. Domain/superfamily/multi-domain hits of model proteins were analysed using NCBI's conserved domain database.

3. Results

Table 1 shows that A0A0B2RQU6, G7L0J5, W9SFC3 and A0A0B0P4Q5 proteins respectively show 85.5%, 84.0%, 84.09% and 82.4% similarities to A0A022RDQ0.

Accession Number	Name of plant	Identity to A0A022RDQ0 protein (%)
A0A0B2RQU6	<i>Glycine soja</i>	85.5
G7L0J5	<i>Medicago truncatula</i>	84.0
W9SFC3	<i>Morus notabilis</i>	84.9
A0A0B0P4Q5	<i>Gossypium arboreum</i>	82.4

Table 1: Sequence identity of proteins to A0A022RDQ0

Table 2 reveals that the four model proteins are ribonucleoproteins that localise in the ribonucleoprotein complex.

Accession Number	Name of plant	Functional Similarity
A0A0B2RQU6	<i>Glycine soja</i>	Ribonucleoprotein
G7L0J5	<i>Medicago truncatula</i>	Ribonucleoprotein
W9SFC3	<i>Morus notabilis</i>	Ribonucleoprotein
A0A0B0P4Q5	<i>Gossypium arboreum</i>	Ribonucleoprotein

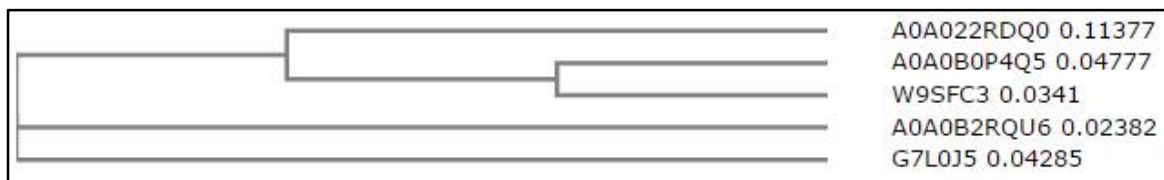
Table 2: Predicted functional similarity in model proteins

Table 3 below shows that the four model proteins localise in the ribonucleoprotein complex.

Accession Number	Name of plant	Subcellular localization
A0A0B2RQU6	<i>Glycine soja</i>	Ribonucleoprotein complex
G7L0J5	<i>Medicago truncatula</i>	Ribonucleoprotein complex
W9SFC3	<i>Morus notabilis</i>	Ribonucleoprotein complex
A0A0B0P4Q5	<i>Gossypium arboreum</i>	Ribonucleoprotein complex

Table 3: Similarity in subcellular localization of proteins as an insight to that of A0A022RDQ0

Phylogenetic tree in figure 1 below report that all proteins have similar evolutionary origin.

Figure 1: Phylogenetic tree of A0A022RDQ0 (*Mimulus guttatus*), A0A0B2RQU6 (*Glycine soja*), G7L0J5 (*Medicago truncatula*), W9SFC3 (*Morus notabilis*) and A0A0B0P4Q5 (*Gossypium arboreum*)

Domain search in table 4 reveals that all proteins belong to WD40.

Accession Number	Name of plant	Domain/superfamily/multi-domain
A0A022RDQ0	<i>Mimulus guttatus</i>	WD40
A0A0B2RQU6	<i>Glycine soja</i>	WD40
G7L0J5	<i>Medicago truncatula</i>	WD40
W9SFC3	<i>Morus notabilis</i>	WD40
A0A0B0P4Q5	<i>Gossypium arboreum</i>	WD40

Table 4: Domain similarity as a confirmation of A0A022RDQ0 protein function

4. Discussion

Sequence identity of corresponding proteins to A0A022RDQ0 report high similarity implying many characteristics including functional properties correspond between these proteins and A0A022RDQ0. This is also an indication that the five plants that include *Mimulus guttatus* belong to the same family which agrees with Sharma and Kaur. (2017) who suggested plants should be categorized based on similar features such as phytochemicals. High sequence similarity among the proteins as an insight to similar function is consistent with the work of Esposti. (2002) who correlated a common function to some proteins on the basis of sequence similarity.

Having confirmed sequence similarity between the various proteins, the predicted function of the model proteins was assessed (Table 2). These originated uniform findings suggesting AOA022RDQ0 functions as a ribonucleoprotein. Corresponding search for subcellular localization of the four model proteins reported they are centred in the ribonucleoprotein complex which agrees with the findings of Filipowicz and Pogacic. (2002) who reported that small nuclear ribonucleoproteins (snRNP) are localized in nucleolus complexes. As a confirmation to the subcellular localization of ribonucleoproteins, Moriguchi et al. (1997) in a western blot experiment analysed RGP-1b protein which is an RNA-binding protein to be highly concentrated in the nucleolus even though there were present in entire nucleus of tobacco cells.

Phylogenetic tree report all the five proteins originate from a common ancestor, Eukaryota. This further informs the reason for high similarity in sequence as reported in table 1 that also account for correlative function which is in accordance with Mount. (2004) who suggested the strongest exhibit for common descent is comparative sequence analysis. Our findings of common evolutionary origin as an insight into similar function is consistent with the work of Petrov et al. (2000) who informed that the configuration of DNA, RNA and amino acids is conserved across generations implying similar proteins would have the same function in different organisms. Diversions in phylogeny could be attributed to junk DNA including pseudogenes that cause mutation thus, reconstructing ancestral lineages as reported by Petrov et al. (2000).

All five proteins belong to WD40 domain which further support results obtained from table 1 and figure 1. This is a clear cut confirmation of the homology of these proteins as well as their distant homology as reported by Lee and Lee. (2009) who compared proteins at domain architecture level using domain-based approaches. High similarity in protein structure and the same pedigree is a confirmation that the five proteins perform similar function.

5. Conclusion

This work is designed at setting the necessary foundation for further studies on AOA022RDQ0 protein. Although establishing our findings with wet laboratory experiments is yet to be carried out because of some limitations, this study reports that AOA0B2RQU6, G7LOJ5, W9SFC3 and AOA0BOP4Q5 are homologues of AOA022RDQ0 protein. AOA022RDQ0 functions as a ribonucleoprotein and is localized in the ribonucleoprotein complex. The five proteins including AOA022RDQ0 originate ancestrally from Eukaryota. AOA022RDQ0 belong to WD40 domain/superfamily/multi-domain.

6. References

- i. Basic Local Alignment Search Tool. (2016). Protein blast. [Accessed 20 January 2016]. Available from: <http://blast.ncbi.nlm.gov>.
- ii. EMBI-EBI. (2016). Clustal Omega Database. [Accessed 21 January 2016]. Available from: <http://www.ebi.ac.uk>.
- iii. Esposti, D.M. (2000). Sequence and functional similarities between pro-apoptotic Bid and plant lipid transfer proteins. *Biochim Biophys Acta*.15(3): 331-40.
- iv. Filipowicz, W and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Current Opinion in Cell Biology*. 14(3):319–327.
- v. Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J and Bateman, A.(2016). *Nucleic Acids Research*. Database Issue 44: 279-285. Available from: <https://pfam.xfam.org>.
- vi. Hellsten, U., Wright, K.M., Jenkins, J., Shu, S., Yuan, Y., Wessler, S.R., Schmutz, J., and
- vii. Rokhsar, D.S. (2013). Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proceedings of the National Academy of Sciences of USA*. 110:19478-19482. Available from: <http://www.uniprot.org>.
- viii. Lee, B and Lee, D. (2009). Protein comparison at the domain architecture level. *BMC Bioinformatics*. 10(15): S5.
- ix. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C and Bryant, S.H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*. 43:222-6. <http://www.ncbi.nlm.nih.gov>.
- x. Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C and Bryant, S.H. (2011). CDD: NCBI's conserved domain database. *Nucleic Acids Research*. 39:225-9. <http://www.ncbi.nlm.nih.gov>.
- xii. Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Lui, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang N and Bryant, S.H. (2009). Specific functional annotation with conserved domain database. CDD: NCBI's conserved domain database. *Nucleic Acids Research*. 37:205-10. <http://www.ncbi.nlm.nih.gov>.
- xiv. Marchler-Bauer, A and Bryant, S.H. (2004). Protein domain annotations on the fly. CD: NCBI's conserved domain database. *Nucleic Acids Research*. 32:327- 331. <http://www.ncbi.nlm.nih.gov>. Moriguchi, K., Sugita, M. and Sugiura, M. (1997). Structure and subcellular localization of a small RNA-binding protein from tobacco. *Plant Journal*. 12(1): 215–221. Mount, D.M. (2004). *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). Cold Spring Harbor Laboratory Press, NY. 600pp.
- xv. Sharma, R and Kaur, S. (2017). Antimicrobial and phytochemical screening of Trikota-traditional food of western Rajasthan. *Indian Journal of Traditional Knowledge*. 16(2): 270-276. Petrov, D.A and Hartl, D.L. (2000). Pseudogene evolution and natural selection for a compact genome. *The Journal of Heredity*. 91(3):221–227.