IJCTA, 9(23), 2016, pp. 183-189

Phylogenetic Analysis and In-Silico Characterization of Nucleotide Binding Site Proteins from Ascochyta Rabiei

Kamran Manzoor Waidha* Apoorv Gupta Rupsneet Kaur*** Pooja Mahapatra**** Divyam Singh***** and Neetu Jabalia********

Abstract : The energy transfer and genetic information are majorly required for Cellular processes. Nucleotide –binding site (NBS) containing proteins play crucial role in defense mechanism in a wide range of species from humans to plants. In this study, computational approaches have been adopted to explore properties of all the 22 NBS sequences from Ascochyta rabiei. Physico-chemical properties have been studies along with the prediction of motifs for the functional characterization. The secondary structure prediction of sequences revealed that the presence of maximum number of random coils probably due to presence of higher amount of more flexible amino acids such as glycine and proline. The phylogenetic studies have been done to find the evolutionary relationship among all the nucleotide binding site proteins.

Keywords : Ascochyta rabiei, nucleotide binding site proteins, motifs.

1. INTRODUCTION

In various cellular processes that occur within the cell ranging from genetic transmission, transfer of energy and storage. Nucleotides play crucial role. The nucleotide triphosphate (NTP) is found in nucleoside monophosphate (NMP) kinases where N can be adenosine or guanosine [1]. Nucleotides consist of three major components such as nitrogenous base, a pentose sugar comprised of 5carbons and phosphate group, constrained by particular structural motifs that repeat in proteins of various fold. These conserved regions (motifs) carry on as modules and are found in various conjunctions crosswise over protein folding. The nucleotide binding site regions are well conserved in numerous organism which is marked by conserved motifs such as kinase 1, kinase 2 and kinase 3 [2].

C-termini of the proteins are highly variable and are marked by the presence of different types of protein-protein interacting domains [3]. In the present paper we investigate the phylogenetic analysis and *in-silico* characterization of 22 nucleotide binding site protein sequences from *Ascochyta rabiei*.

2. MATERIALS AND METHOD

2.1. Sequence retrieval

The nucleotide binding site protein sequences of *Ascochyta rabiei* were searched and obtained from NCBI database using the search query "nucleotide binding protein and Ascochyta rabiei", total 22 sequence were found.

* Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida
** Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida
*** Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida
**** Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida
***** Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida
****** Amity Institute of Biotechnology, Amity University Uttar Pradesh, Noida

To verify the presence of nucleotide binding site protein domains, protein sequences were searched against the InterProScan database [4].

2.2. Functional characterization

The presence of the conserved motifs present within protein sequences of nucleotide binding site (NBS) domains were scanned using Motif Search. For analysis of hydrophobicity of sequences SOSUI server was used [5].

2.3. Prediction of Primary Structure using physio- chemical properties

Various physical and chemical properties of sequences were calculated by Protparam Tool provided by EXpasy, properties computed included calculation of isoelectric point(pI), estimation of molecular weight, instability index [6], extinction coefficient [7] aliphatic index [8] and GRAVY [9][10] .

Secondary structure prediction : The analysis of the secondary structure of nucleotide binding site proteins was done by using GORIV. The secondary structure prediction is based on the information of their primary structure prediction.

2.4. Phylogenetic Analysis

For alignment of multiple sequences, the obtained 22 nucleotide binding site protein sequences were aligned by the ClustalW [11] with default options. On the basis of neighbor-joining method a phylogenetic tree was constructed[12] using the JTT model with the help of a MEGA 7 software [13].The test of phylogeny and determination of internal nodes stability was done by bootstrap method for 3000 replications.

3. RESULT AND DISCUSSION

The nucleotide binding site proteins (NBS) sequences from *Ascochyta rabiei* were retrieved from NCBI. All the sequence were found to be playing vital role in nucleotide binding site. A set of conserved amino acid residues located in the vicinity that provide clues to the functions is termed as motif. Motif were predicted using MotifSearch (Table 1). It has been predicted that out of total 22 NBS proteins all have RNA recoginition motifs (RRM) except KZM26879.1, KZM26565.1, KZM24134.1, KZM23927.1, KZM23842.1, KZM22992.1, KZM21851.1, KZM21809.1 and KZM21461.1.

Table 1. Motifs prediction of NBS from *Ascochyta rabiei*

S.No.	*Accession No.*	*Motif Description*	*Start*	*End*
1.	KZM28493.1	RRM_1	49	118
			137	204
			230	298
			333	365
			434	434
		RRM_5	30	125
			135	207
		RRM_occluded	232	304
2.	KZM28245.1	RRM_1	203	269
		RRM_3	208	262
		RRM_5	203	279
		RRM_occluded	201	273

S.No.	*Accession No.*	*Motif Description*	*Start*	*End*
3.	KZM28016.1	RRM_1	285	354
			400	464
			528	588
		RRM_occluded	398	431
		Nup35_RRM_2	538	574
4.	KZM27781.1	RRM_1	114	128
5.	KZM27158.1	RRM_1	53	85
6.	KZM26212.1	RRM_1	172	239
			274	345
			500	545
		RBM39linker	411	482
		RRM_occluded	282	348
			497	549
		Nup35_RRM_2	186	230
		RRM_5	283	358
7.	KZM26182.1	RRM_1	48	118
8.	KZM23231.1	RRM_1	76	117
9.	KZM23195.1	RRM_1	24	93
		RRM_5	20	103
10.	KZM21188.1	RRM_1	118	154

SOSUI server was used to characterize whether the proteins would be soluble or transmembrane in nature. All the nucleotide binding site proteins of *Ascochyta rabiei* were found to be soluble (Table 2).

Table 2. Prediction of hydropathicity for the NBS protein sequence from *Ascochyta rabiei*

S. No.	*Accession number*	*Protein Description*	*Average of Hydrophobicity*
1.	KZM28493.1	Soluble Protein	-0.799204
2.	KZM28245.1	Soluble Protein	-0.8175
3.	KZM28016.1	Soluble Protein	-0.968824
4.	KZM27781.1	Soluble Protein	-0.969868
5.	KZM27158.1	Soluble Protein	-1.220102
6.	KZM26879.1	Soluble Protein	-0.958955
7.	KZM26565.1	Soluble Protein	-0.998775
8.	KZM26212.1	Soluble Protein	-1.081627
9.	KZM26182.1	Soluble Protein	-1.090964
10.	KZM24134.1	Soluble Protein	-0.664238
11.	KZM23927.1	Soluble Protein	-0.380940
12.	KZM23842.1	Soluble Protein	-0.145361

S. No.	*Accession number*	*Protein Description*	*Average of Hydrophobicity*
13.	KZM23231.1	Soluble Protein	-0.614054
14.	KZM23195.1	Soluble Protein	-0.76776
15.	KZM22992.1	Soluble Protein	-0.557244
16.	KZM22141.1	Soluble Protein	-0.864923
17.	KZM21851.1	Soluble Protein	-0.339216
18.	KZM21809.1	Soluble Protein	-0.568315
19.	KZM21461.1	Soluble Protein	-0.41226
20.	KZM21188.1	Soluble Protein	-1.120678
21.	KZM19698.1	Soluble Protein	-1.084577
22.	KZM19410.1	Soluble Protein	-0.432793

The values of isoelectric point (pI) of the nucleotide binding site proteins were in the range 4.42 to 10.75. indicating that KZM28245.1, KZM27158.1, KZM26879.1, KZM26565.1, KZM26212.1, KZM22141.1, KZM21851.1, KZM21188.1 and KZM19410.1 are basic while others are acidic proteins (Table 3). The isoelectric point will be beneficial for preparing buffer system when these proteins are to be purified in solution by isoelectric focusing method [14]. The Protparam was used to determine the extinction coefficient of protein at 280nm ranges from 4470 to 146915 M^{-1} cm^{-1} with respect to the concentration of Cys, Trp and Tyr will be useful in the quantitative study of protein-protein and protein-ligand interactions in solution. The aliphatic index measures the relative volume of a protein occupied by aliphatic side chains, as regarded as a positive factor for the increase of thermal stability of globular proteins. Aliphatic index for NBS from *Ascochyta rabiei* ranged from 27.79 to 88.24. The very high aliphatic index of the sequences indicate that proteins will be stable over a wide temperature range [15]. The instability index provides an estimate of the stability of our protein. It has been found that all the nucleotide binding site proteins were unstable except KZM26879.1, KZM21851.1 and KZM19410.1 were stable. The grand average hydropathicity value (GRAVY) for NBS proteins are calculated as the sum of hydropathy values of all the amino acids, divided by the total number of residues present in nucleotide binding site proteins [16]. GRAVY indices for the sequences ranged from -1.121 to -0.339. This low range values indicated better interaction with water [17].

Table 3. Physico-chemical properties of NBS proteins from *Ascochyta rabieis*

Accesion No.	*Molecular weight*	*Theoretical pI*	*The instability index (II)*	*Grand average of hydropathicity (GRAVY)*	*Ext. coefficient*	*Aliphatic index*
KZM28493.1	82281.3	5.61	45.99	-0.799	35550	55.67
KZM28245.1	30783.7	8.98	41.3	-0.818	22710	54.43
KZM28016.1	66336.7	5.63	55.48	-0.969	43320	59.6
KZM27781.1	25231.6	4.8	47.14	-0.97	9970	56.29
KZM27158.1	65989.5	8.99	57.31	-1.22	47440	51.28
KZM26879.1	15392.9	10.75	23	-0.959	5960	61.87
KZM26565.1	46176	9.31	47.76	-0.999	32430	60.02
KZM26212.1	63052.3	9.39	48.76	-1.082	38975	56.06
KZM26182.1	19224.3	5.56	42.19	-1.091	18130	52.29
KZM24134.1	33538.1	5.14	54.96	-0.664	35535	64.9
KZM23927.1	43218.8	4.47	53.75	-0.381	37150	87.57

Accesion No.	*Molecular weight*	*Theoretical pI*	*The instability index (II)*	*Grand average of hydropathicity (GRAVY)*	*Ext. coefficient*	*Aliphatic index*
KZM23842.1	32705.4	4.42	52.68	-0.145	12170	88.14
KZM23231.1	84399.7	5.06	42.83	-0.614	146915	77.96
KZM23195.1	20267.7	4.84	51.38	-0.768	5960	61.97
KZM22992.1	46084	4.61	52.9	-0.557	22140	82.59
KZM22141.1	36425.8	9.86	66.95	-0.865	23950	67.88
KZM21851.1	16624.4	10.43	33.78	-0.339	4470	88.24
KZM21809.1	49241.1	4.85	53.35	-0.568	27515	64.45
KZM21461.1	28733.6	5.44	58.48	-0.412	16960	77.74
KZM21188.1	36890	7.83	48.26	-1.121	33460	63.55
KZM19698.1	42533	5.59	36.63	-1.085	23950	27.79
KZM19410.1	27589.6	6.87	40.43	-0.433	10430	72.31

Secondary structure features were predicted by GORIV (Figure 1) The results revealed that out of 22 nucleotide binding site protein sequences random coils dominated among all the secondary structure elements except KZM23927.1, KZM23842.1, KZM23195.1, KZM22992.1, KZM22141.1, KZM21851.1, KZM21188.1 and KZM19410.1 having high helix content. The conformational entropy associated with random coils significantly contributes to stabilization and protein folding.

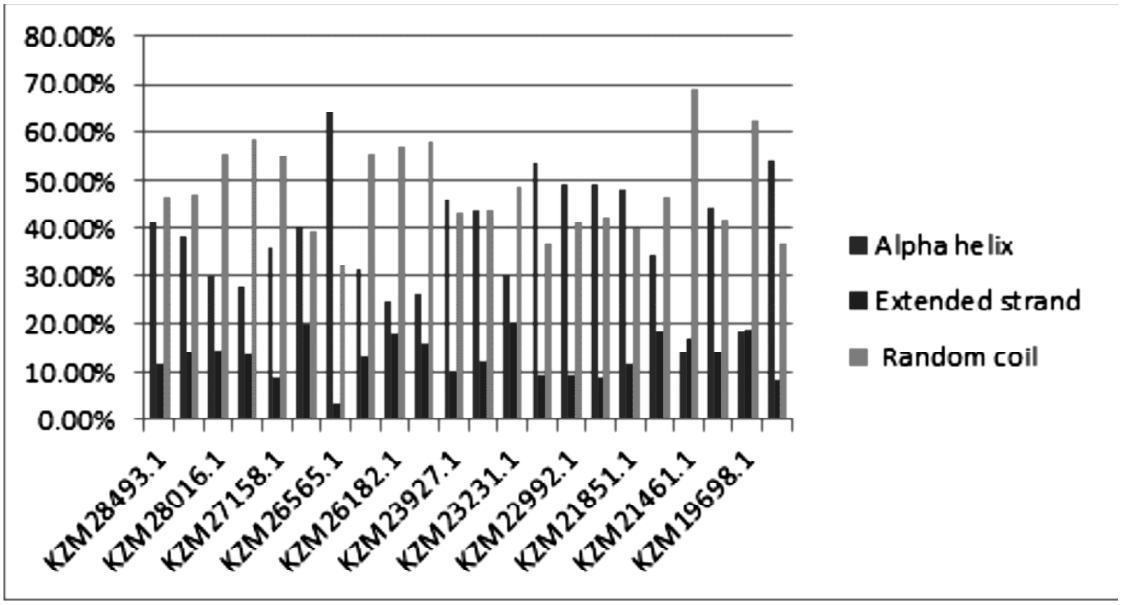

Fig. 1. Graphical representation of the content of secondary structure elements of the NBS protein from *Ascochyta rabiei*

Proline, which has a high content in the nucleotide binding site proteins, has special property of creating kinks in polypeptide chains and disrupting ordered secondary structure and might have contributed to high content of random coils structure.

To clarify the phylogenetic relationship among the NBS proteins and infer the evolutionary history of the family, a phylogenetic tree was constructed using the protein sequences of the NBS region. This tree was generated from neighbor joining method based on multiple sequence alignments of the nucleotide binding site proteins sequences done by MEGA 7.0. Figure 2 showed almost similar distribution of the *Ascochyta rabiei* nucleotide binding site sequences.

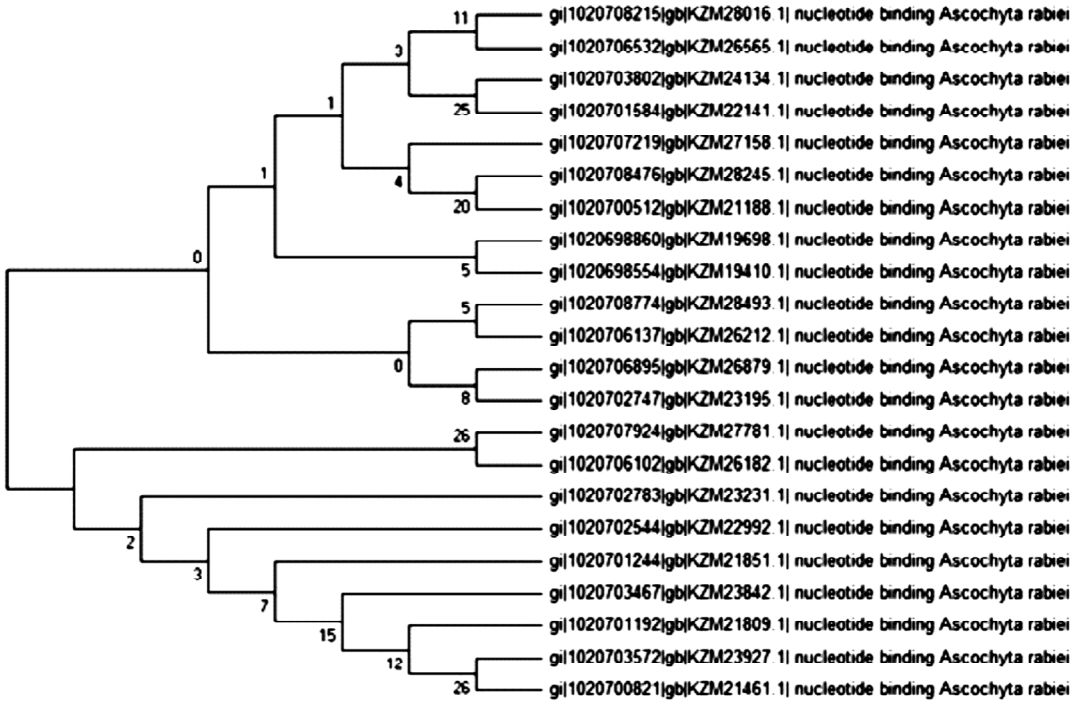

Fig. 2. Phylogenic tree of the *Ascochyta rabiei* NBS proteins sequences based on the amino sequences in the NBS domain

4. CONCLUSION

In our investigation, we have characterized 22 nucleotide binding site protein sequences to acquire acknowledge about their functional properties, physical as well as chemical properties of different protein structure levels by using computational approaches. Proteins with NBS domain are playing vital role in programmed cell death in metazoans and disease resistance in plants. Primary structure analysis revealed that nucleotide binding site proteins from *Ascochyta rabiei* were hydrophilic and are expected to be stable over wide range of temperature. Secondary structure analysis established that in most of the sequences, random coils were the dominating secondary structure elements followed by alpha helix, extended stand and beta turns. Phylogenetic analysis revealed that having nucleotide binding site domain with similar distribution. among all the sequences of *Ascochyta rabiei*. The future research can explain functional, evolutionary and structural importance of these protein domains

5. REFERENCES

1. L. Parca, P. F. Gherardini, M. Truglio, I. Mangone, F. Ferrè, M. Helmer-Citterich., A. Gabriele, "Identification of Nucleotide-Binding Sites in Protein Structures: A Novel Approach Based on Nucleotide Modularity", *PLoS ONE 7(11) (2012) e50240*.

2. E. A. Biezen., J. D. Jones, "The NB-ARC domain: A novel signaling motif shared by plant resistance gene product and regulators of cell death in animals", *Currrent Biology 8 (1998) 226-227*.

3. J. K. Dangl, J. D. G. Jones, "Plants pathogens and integrated defence responses to infection", *Nature* 411 (2001) 826-833.

4. E. Quevillon, V. Silventoninen, S. Pillai, Harte N., N. Mulder, "InterProScan: protein domains identifier", *Nucleic Acids Research* 33 (2005) 116-120.

5. T. Hirokawa, "SOSUI: classification and secondary structure prediction system for membrane proteins", *Bioinformatics*, 14 (1998) 378-379.

6. K. Guruprasad, "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence", *Protein Eng. 4*, (1990) 155-161.

7. S. Gill, C. Von, P. H. Hippel, "Calculation of protein extinction coefficients from amino acid sequence data", *Anal. Biochem.*182, 319-326.

8. A.J. Ikai, "Thermostability and aliphatic index of globular proteins" *J. Biochem.* 88, 1895-1898.

9. J. Kyte, R.F. Doolittle, "A simple method for displaying the hydropathic character of a protein" *J. Mol. Biol. 157*, 105-132 (1982).

10. E . Gasteiger, and M. Walker John, (ed). *Protein Identification and Analysis Tools on the ExPASy Server,* The Proteomics Protocols Handbook, Humana Press. 571-607, 2005.

11. J.D. Thompson , T.J. Gibson , D.G. Higgins " Multiple sequence alignment using ClustalW and ClustalX" *Curr Protoc Bioinformatics*. Chapter 2. August, 2002.

12. N. Saitou, M. Nei,. "The neigbour-joining method: a new method for reconstructing phylogentic trees". *Mol Biol Evol* (4) 406-425. 1987

13. S. Kumar , G. Stecher , K. Tamura "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets". *Molecular biology and evolution*, 2016.

14. N. K. Verma, B. Singh, "Insight from the structural molecular model of cytidylate kinase from *Mycobacterium tuberculosis"*. *Bioinformation (9)* 680–684, 2013.

15. H. Bansal, S. Srivastava, A. Chaurasia, N. Jabalia,. "A Comparative Study of Antifreeze Proteins from Antarctomyces psychrotrophicus and Typhula ishikariensis using Computational Tools and Servers" *VIVECHAN Int. J. Res., 5* 21-28, 2014.

16. H. Bansal, D. Narang, N. Jabalia "Computational characterization of antifreeze proteins of Typhula ishikariensis – Gray Snow Mould". *J. Proteins Proteomics* 5, 169-179, 2004.

17. N. Jabalia, H. Bansal, P.C. Mishra, N. Chaudhary, "*In-silico* comparative analysis of papain family cysteine protease using computational tools and servers". *Int. J. Basic and Appl. Eng*. Res. 2, 310-314, 2015.