ECRONICON OPEN ACCESS

EC PROTEOMICS AND BIOINFORMATICS
Research Article

Functionality Search for Structure and Sub-Cellular Localization of Hypothetical Proteins in Haloarchaeon *Natrinema* sp. J7-2 Plasmid

Hitesh S Thakare[1,2]*, Satish Sawale[1], Dilip B Meshram[3], Kunal Roychoudhary[2] and Arun B Ingle[2]

[1]National Environmental Engineering Research Institute, Nagpur, Maharashtra, India

[2]Department of Microbiology, Seth Kesarimal Porwal College, Maharashtra, India

3Department of Zoology, Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur, Maharashtra, India

***Corresponding Author:** Hitesh S Thakare, National Environmental Engineering Research Institute, Nagpur and Department of Microbiology, Seth Kesarimal Porwal College, Maharashtra, India.

Received: August 03, 2018; **Publisded:** October 30, 2019

Abstract

The hypothetical proteins in *Natrinema* sp. J7-2 plasmid were explained based on their structure, function and sub-cellular localization. The total of 72 genes was screened for structures, functions and sub-cellular localization of hypothetical proteins in *Natrinema* sp. J7-2 plasmid out of these 48 genes was predicted for hypothetical proteins which were unknown for their structures, functions and sub-cellular localization. The CDD-BLAST, INTERPROSCAN and PFAM were used to determine the functional annotations. The sub- cellular localizations and presence of templates for conserved domains were determined by PS2 Server-Protein Structure Prediction server respectively. The PS2 Server also used to determine 3-D structures, E-value and aligned percentage of the predicted proteins. The present study may be useful for understanding the genetics and metabolic pathways at the molecular level as well as functional characteristics of hypothetical proteins in ecologically important haloarchaeon *Natrinema* Sp. J7-2.

Keywords: *Haloarchaeon; Glycolysis; E-value; Templates; INTERPROSCAN*

Introduction

Natrinema sp. J7-2 is an extreme haloarchaeon. It was first isolated from a salt mine in China [1]. Haloarchaeon requires hypersaline environment for growth and this type of environments are commonly found in China [2]. By culture, dependent molecular methods various types of Halophilic archaeon found in food products also identified in high salt-fermented foods consumed by humans [3]. Halophilic archaeon was non-motile, neutrophilic, coccus or irregularly shaped and rod [2].

Haloarchaeon is an ideal to explain the metabolic pathways since it grows in synthetic media containing no any other supplements such as amino acids and vitamins. It uses gluconate, glycerol or acetate as carbon source. It cannot follow the classical glycolysis pathway since the enzyme 6-phosphofructokinase is absent [1].

Most of Haloarchaeon provide osmotic balance by using organic solutes in between cytoplasm and the surrounding medium. They have many applications in the bioremediation, biomaterial and nanotechnology research areas also they have been focused on environmental and ecological researchers [3,4]. Haloarchaea requires 1.5M to 2.5M salt concentration for growth therefore, they are referred to as extremely halophilic archaea or haloarchaea [3].

Studies of the genus *Natrinema* have mainly focused on the phenotypic, physiological, biochemical and taxonomic properties of genus *Natrinema* have been studied very well but no data available on their genetics. The study carried out on the genomes of only two strains (J7-1 and J7-2) revealed that they were genetically identical [5,6].

The genome of *Natrinema Sp.* J7-2 plasmid contains 72 predicted protein-coding genes. A total of 48 conserved hypothetical genes. In the sequence databases, there are no significant matches of these genes. The bioinformatics tools and servers were used to reveal the function of particular genes which have the ability to determine the presence of the enzymatic conserved domain/s. It also helps in a prediction of the three-dimensional structures. CDD-BLAST, INTERPROSCAN, and PFAM are used to understand the function of the protein sequence. The sub-cellular localization of protein or enzyme was determined using Cello. The Protein Structure Prediction Server (PS2 server) can be used for prediction of the structure of proteins.

Aim of the Study

The present work aimed to understand the structure and functions of hypothetical proteins in Plasmid *Natrinema Sp.* J7-2 by using online bioinformatics tools and servers.

Methodology

Sequence Retrieval

The KEGG database was used for retrieval of *Natrinema Sp.* J7-2 plasmid whole genome sequences (http://www.genome.jp/kegg/).

Functional annotation and categorization

The presence of conserved functional domains and 3-D structures of the hypothetical proteins from *Natrinema sp.* J7-2 Plasmid were screened and analyzed by using the bioinformatics web tools. The CDD-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/; [7-10]){ “id” : “ITEM-2”, “itemData” : { “DOI” : “10.1093/nar/29.14.2994”, “ISBN” : “0305-1048”, “ISSN” : “1362-4962”, “PMID” : “11452024”, “abstract” : “PSI-BLAST is an iterative program to search a database for proteins with distant similarity to a query sequence. We investigated over a dozen modifications to the methods used in PSI-BLAST, with the goal of improving accuracy in finding true positive matches. To evaluate performance we used a set of 103 queries for which the true positives in yeast had been annotated by human experts, and a popular measure of retrieval accuracy (ROC, INTERPROSCAN [11])”id” : “ITEM-1”, “issued” : { “date-parts” : [[“0”]] }, “title” : “No Title”, “type” : “webpage” }, “uris” : [“http://www.mendeley.com/documents/?uuid=60623971-5f36-48ae-b749-d6837a17dd0d”] }, { “id” : “ITEM-2”, “itemData” : { “DOI” : “10.1093/bioinformatics/17.9.847”, “ISBN” : “1367-4803 (Print, Pfam (http://www.pfam.sanger.ac.uk/; [12]) and Cello were used as bioinformatics web tools (http://cello.life.nctu.edu.tw/). CDD-Blast, Interproscan, and Pfam were used for the presence of conserved domains and functional characteristics in the sequences as per the information available in databases. The Cello server v 2.5 was used to determine the sub-cellular localization of the identified protein or enzyme within the cell.

Protein structure prediction

Protein Structure Prediction Server (PS2 server) is used for the prediction of the 3-D structure of hypothetical proteins (http://www.ps2.life.nctu.edu.tw/; [10,13])we illustrate the power of the approach by using a combination of local and global pair-wise alignments to generate the library. The resulting alignments are significantly more reliable, as determined by comparison with a set of 141 test cases, than any of the popular alternatives that we tried. The improvement, especially clear with the more difficult test cases, is always visible, regardless of the phylogenetic spread of the sequences in the tests.”, “author” : [{ “dropping-particle” : “”, “family” : “Notredame”, “given” : “C”, “non-dropping-particle” : “”, “parse-names” : false, “suffix” : “” }, { “dropping-particle” : “”, “family” : “Higgins”, “given” : “D G”, “non-dropping-particle” : “”, “parse-names” : false, “suffix” : “” }, { “dropping-particle” : “”, “family” : “Heringa”, “given” : “J”, “non-dropping-particle” : “”, “parse-names” : false, “suffix” : “” }], “container-title” : “Journal of molecular biology”, “id” : “ITEM-2”, “issue” : “1”, “issued” : { “date-parts” : [[“2000”]] }, “page” : “205-217”, “title” : “T-Coffee: A novel method for fast and accurate multiple sequence alignment.”, “type” : “article-journal”, “volume” : “302” }, “uris” : [“http://www.mendeley.com/documents/?uuid=0dbaf73e-a47f-4424-8a61-617868fe9969”] }, { “id” : “ITEM-3”, “itemData” : { “DOI” : “10.1186/1471-2105-7-178”, “ISBN” : “1471-2105”, “ISSN” : “1471-2105”, “PMID” : “16571137”, “abstract” : “The accuracy of protein secondary structure prediction has been improving steadily towards the 88% estimated theoretical

limit. There are two types of prediction algorithms: Single-sequence prediction algorithms imply that information about other (homologous. The prediction of protein 3D structures generated while running the FASTA format of protein sequence by using this online web server. The structural model of the protein is based on the functional annotations for template detection.

Results and Discussion

The computational studies were carried out for characterizing 48 hypothetical proteins from the complete genome sequences for *Natrinema sp.* J7-2 Plasmid. CDD- Blast, Interproscan, Pfam, Cello and PS2server (available online web servers) were used for predictions of structural, and functional characteristics of total 48 hypothetical proteins. Sub-cellular localization of all these hypothetical proteins was characterized successfully. The scoring templates for 3-D structures are represented in the order as Template ID, E-value and aligned percentage in PS2 structure template column (Table 1).

NCBI gene ID	Functionality of unknown proteins			Sub-Cellular localization	3-D structures prediction for conserved domains and templates		
	CDD Blast	Interproscan	Pfam	Cello V2.5	PS2 Server	E-value	Aligned percentage
13353376	NA	NA	NA	Cytoplasmic 4.434	NA	NA	NA
13353377	NA	NA	NA	Cytoplasmic 3.348	3etfA	5.5	94.52
13353378	NA	NA	NA	Cytoplasmic 3.075	2f23A	1.3	76.24
13353379	NA	NA	NA	Cytoplasmic 2.454	NA	NA	NA
13353380	NA	NA	NA	Cytoplasmic 3.202	1x4xA	6.2	64.86
13353381	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors. ARSR subfamily of helix-turn-helix bacterial transcription regulatory proteins (winged helix topology). Includes several proteins that appear to dissociate from DNA in the presence of metal ions.	Sugar-specific transcriptional regulator TrmB,N-terminal, Winged helix-turn-helix DNA-binding domain	Sugar-specific transcriptional regulator TrmB	Cytoplasmic 4.141	2p4wB	0.02	83.67
13353382	NA	NA	NA	Cytoplasmic 2.973	NA	NA	NA
13353383	SoSSB_OBF: A subfamily of OB folds similar to the OB fold of the crenarchaeote Sulfolobus solfataricus single-stranded (ss) DNA-binding protein (SSoSSB). SSoSSB has a single OB fold, and it physically and functionally interacts with RNA polymerase. In vitro, SSoSSB can substitute for the basal transcription factor TBP, stimulating transcription from promoters under conditions in which TBP is limiting, and supporting transcription when TBP is absent. SSoSSB selectively melts the duplex DNA of promoter sequences. It also relieves transcriptional repression by the chromatin Alba. In addition, SSoSSB activates reverse gyrase activity, which involves DNA binding, DNA cleavage, strand passage and ligation. SSoSSB stimulates all these steps in the presence of the chromatin protein, Sul7d. SSoSSB antagonizes the inhibitory effect of Sul7d on reverse gyrase supercoiling activity. It also physically and functionally interacts with Mini-chromosome Maintenance (MCM), stimulating the DNA helicase activity of MCM.	SUPERFAMILY SSF50249, Nucleic acid-binding, OB-fold	NA	Cytoplasmic 4.269	2k75A	0.0044	35.07
13353385	NA	NA	NA	Cytoplasmic 2.812	2oikC	0.89	47.44
13353386	NA	NA	NA	Cytoplasmic 3.040	NA	NA	NA
13353387	NA	NA	NA	Cytoplasmic 4.246	1zovA	8.1	77.60
13353388	NA	NA	NA	Cytoplasmic 2.869	1ei1A	8.4	100.00
13353389	Domain of unknown function (DUF3883); This is a domain is uncharacterized. It is found on restriction endonucleases.	Domain of unknown function (DUF3883), Histidine kinase-like ATPase, C-terminal domain	Domain of unknown function (DUF3883)	Cytoplasmic 2.814	NA	NA	NA
13353390	NA	NA	NA	Cytoplasmic 3.498	1egzA	8.9	73.19
13353391	NA	Coils Coil	NA	Cytoplasmic 4.121	1oxxK	4.8	95.71
13353392	NA	NA	NA	Periplasmic 2.066 Cytoplasmic 1.392	2jdcA	5.9	85.12

13353393	ZPR1 zinc-finger domain; The zinc-finger protein ZPR1 is ubiquitous among eukaryotes. It is indeed known to be an essential protein in yeast. In quiescent cells, ZPR1 is localized to the cytoplasm. But in proliferating cells treated with EGF or with other mitogens, ZPR1 accumulates in the nucleolus. ZPR1 interacts with the cytoplasmic domain of the inactive EGF receptor (EGFR) and is thought to inhibit the basal protein tyrosine kinase activity of EGFR. This interaction is disrupted when cells are treated with EGF, though by themselves, inactive EGFRs are not sufficient to sequester ZPR1 to the cytoplasm. Upon stimulation by EGF, ZPR1 directly binds the eukaryotic translation elongation factor-1alpha (eEF-1alpha) to form ZPR1/eEF-1alpha complexes. These move into the nucleus, localizing particularly at the nucleolus. Indeed, the interaction between ZPR1 and eEF-1alpha has been shown to be essential for normal cellular proliferation, and ZPR1 is thought to be involved in pre-ribosomal RNA expression. The ZPR1 domain consists of an elongation initiation factor 2-like zinc finger and a double-stranded beta helix with a helical hairpin insertion. ZPR1 binds preferentially to GDP-bound eEF1A but does not directly influence the kinetics of nucleotide exchange or GTP hydrolysis. The alignment for this family shows a domain of which there are two copies in ZPR1 proteins. This family also includes several hypothetical archaeal proteins (from both Crenarchaeota and Euryarchaeota), which only contain one copy of the aligned region. This similarity between ZPR1 and archaeal proteins was not previously noted.	NA	NA	Cytoplasmic 3.712	NA	NA	NA
13353394	Domain of unknown function (DUF4209); This short domain is found in bacteria and eukaryotes, though not in yeasts or Archaea. It carries a highly conserved RNxxxHG sequence motif.	Phobius NON_CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region, Phobius CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm,Phobius TRANSMEMBRANE Region of a membrane-bound protein predicted to be embedded in the membrane, Coils Coil, Domain of unknown function DUF4209	Domain of unknown function (DUF4209)	Cytoplasmic 4.785	2qnaA	0.49	98.32
13353395	NA	NA	NA	Cytoplasmic 4.672	1ugpA	1.5	98.42
13353397	NA	NA	NA	Cytoplasmic 3.559	2nsmA	0.11	85.53
13353398	NA	Coils Coil	NA	Cytoplasmic 2.792	NA	NA	NA
13353399	TIM barrel proteins share a structurally conserved phosphate binding motif and in general share an eight beta/alpha closed barrel structure. Specific for this family is the conserved phosphate binding site at the edges of strands 7 and 8. The phosphate comes either from the substrate, as in the case of inosine monophosphate dehydrogenase (IMPDH), or from ribulose-5-phosphate 3-epimerase (RPE) or from cofactors, like FMN.	Coils Coil	NA	Cytoplasmic 4.681	NA	NA	NA
13353401	Domain of unknown function (DUF1837); This family of proteins are functionally uncharacterized.	Protein of unknown function DUF1837	Domain of unknown function (DUF1837)	Cytoplasmic 2.543			
13353402	GIY-YIG nuclease domain superfamily; The GIY-YIG nuclease domain superfamily includes a large and diverse group of proteins involved in many cellular processes, such as class I homing GIY-YIG family endonucleases, prokaryotic nucleotide excision repair proteins UvrC and Cho, type II restriction enzymes, the endonuclease/reverse transcriptase of eukaryotic retro transposable elements, and a family of eukaryotic enzymes that repair stalled replication forks. All of these members contain a conserved GIY-YIG nuclease domain that may serve as a scaffold for the coordination of a divalent metal ion required for catalysis of the phosphodiester bond cleavage. By combining with different specificity, targeting, or other domains, the GIY-YIG nucleases may perform different functions.	NA	NA	Cytoplasmic 3.383	NA	NA	NA
13353403	NA	NA	NA	Cytoplasmic 2.641	2q07A	2.9	91.18
13353404	NA	NA	NA	Cytoplasmic 3.640	1xviB	7.9	58.26
13353405	Domain of unknown function (DUF955); Family of bacterial and viral proteins with undetermined function. A conserved H-E-X-X-H motif is suggestive of a catalytic active site and shows similarity to pfam01435.	ProSite Profiles PS50972 Pterin-binding domain profile.	NA	Cytoplasmic 4.224	1gc5A	1.1	95.91

13353407	Arsenical Resistance Operon Repressor and similar prokaryotic, metal regulated homodimeric repressors. ARSR subfamily of helix-turn-helix bacterial transcription regulatory proteins (winged helix topology). Includes several proteins that appear to dissociate from DNA in the presence of metal ions, Helix-turn-helix XRE-family like proteins. Prokaryotic DNA binding proteins belonging to the xenobiotic response element family of transcriptional regulators.	Winged helix-turn-helix DNA-binding domain, SUPERFAMILY, PF12802 MarR family	MarR family	Cytoplasmic 4.819	NA	NA	NA
13353408	NA	Coils Coil	NA	Cytoplasmic 2.807	3bg3B	3.9	85.05
13353411	NA	Coils Coil	NA	Cytoplasmic 3.138	NA	NA	NA
13353412	Protein of unknown function DUF262;Protein of unknown function (DUF1524); This family of uncharacterized proteins contain a conserved HXXP motif. A similar motif is seen in protein families in the His-Me finger endonuclease superfamily which suggests this family of proteins may also act as endonucleases.	Domain of unknown function DUF262, Domain of unknown function DUF1524	Protein of unknown function DUF262, Protein of unknown function (DUF1524)	Cytoplasmic 3.240	2pftA	4.4	69.23
13353413	Uncharacterized conserved protein [Function unknown]	Protein of unknown function DUF159, Uncharacterised ACR, COG2135	NA	Cytoplasmic 3.802	2icuB	12	85.94
13353415	Methyl transferase domain; This family contains methyltransferase domains, N-terminal catalytic domain of the Serine/Threonine Kinase, Mitogen and stress-activated kinase 2; STKs catalyze the transfer of the gamma-phosphoryl group from ATP to serine/threonine residues on protein substrates. MSK2 and MSK1 play nonredundant roles in activating histone H3 kinases, which play pivotal roles in compaction of the chromatin fiber. MSK2 is the required H3 kinase in response to stress stimuli and activation of the p38 MAPK pathway. MSK2 also plays a role in the pathogenesis of psoriasis. MSKs contain an N-terminal kinase domain (NTD) from the AGC family and a C-terminal kinase domain (CTD) from the CAMK family, similar to 90 kDa ribosomal protein S6 kinases (RSKs). MSKs are activated by two major signaling cascades, the Ras-MAPK and p38 stress kinase pathways, which trigger phosphorylation in the activation loop (A-loop) of the CTD of MSK. The active CTD phosphorylates the hydrophobic motif (HM) of NTD, which facilitates the phosphorylation of the A-loop and activates the NTD, which in turn phosphorylates downstream targets. The MSK2 subfamily is part of a larger superfamily that includes the catalytic domains of other protein STKs, protein tyrosine kinases, RIO kinases, aminoglycoside phosphotransferase, choline kinase, and phosphoinositide 3-kinase, type II restriction m6 adenine DNA methyltransferase, Alw26I/Eco31I/Esp3I family; Members of this family are the m6-adenine DNA methyltransferase protein, or domain of a fusion protein that also carries m5 cytosine methyltransferase activity, of type II restriction systems of the Alw26I/Eco31I/Esp3I family. A methyltransferase of this family is alway accompanied by a type II restriction endonuclease from the Alw26I/Eco31I/Esp3I family (TIGR02986) and by an adenine-specific modification methyltransferase. Members of this family are unusual in that regions of similarity to homologs outside this family are circularly permuted. [DNA metabolism, Restriction/modification], Type I restriction-modification system methyltransferase subunit [Defense mechanisms]	DNA methylase, N-6 adenine-specific, conserved site, S-adenosyl-L-methionine-dependent methyltransferase, N12 class	Methyltransferase domain	Cytoplasmic 3.141	1tr2B	0.13	9.73
13353416	NA	Phobius NON_CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region, TMHMM TMhelix Region of a membrane-bound protein predicted to be embedded in the membrane, Phobius TRANSMEMBRANE Region of a membrane-bound protein predicted to be embedded in the membrane.	NA	Cytoplasmic 3.589	NA	NA	NA
13353417	ATPase involved in DNA replication initiation [DNA replication, recombination, and repair]	P-loop containing nucleoside triphosphate hydrolase , Coils Coil	NA	Cytoplasmic 2.975	1tr2B	0.0043	92.58

13353418	Domain of unknown function (DUF1788); Putative uncharacterized domain in proteins of length around 200 amino acids.	Protein of unknown function DUF1788, Coils Coil	Domain of unknown function (DUF1788)	Cytoplasmic 4.468	NA	NA	NA
13353419	NA	Coils Coil, P-loop containing nucleoside triphosphate hydrolase,	NA	Cytoplasmic 3.136	1tr2B	0.002	87.82
13353420	Putative inner membrane protein (DUF1819); These proteins are functionally uncharacterized. Several are annotated as putative inner membrane proteins.	Putative inner membrane protein (DUF1819)	Putative inner membrane protein (DUF1819)	Cytoplasmic 3.113	NA	NA	NA
13353424	NA	NA	NA	Cytoplasmic 2.755	1eq1A	7.6	98.52
13353427	Competence protein CoiA-like family; Many of the members of this family are described as transcription factors. CoiA falls within a competence-specific operon in Streptococcus. CoiA is an uncharacterized protein.	Competence protein CoiA-like family	Competence protein CoiA-like family	Cytoplasmic 4.037	1vq8B	1.2	79.31
13353428	Histone 2A; H2A is a subunit of the nucleosome. The nucleosome is an octamer containing two H2A, H2B, H3, and H4 subunits. The H2A subunit performs essential roles in maintaining structural integrity of the nucleosome, chromatin condensation, and binding of specific chromatin-associated proteins.	Histone-fold, Coils Coil	NA	Cytoplasmic 2.785	NA	NA	NA
13353431	SirA, YedF, and YeeD. Two-layered alpha/beta sandwich domain. SirA (also known as UvrY, and YhhP) belongs to a family of bacterial two-component response regulators that controls secondary metabolism and virulence. The other member of this two-component system is a sensor kinase called BarA which phosphorylates SirA. A variety of microorganisms have similar proteins, all of which contain a common CPxP sequence motif in the N-terminal region. YhhP is suggested to be important for normal cell division and growth in rich nutrient medium. Moreover, despite a low primary sequence similarity, the YccP structure closely resembles the non-homologous C-terminal RNA-binding domain of E. coli translation initiation factor IF3. The signature CPxP motif serves to stabilize the N-terminal helix as part of the N-capping box and might be important in mRNA-binding.	Coils Coil	NA	Cytoplasmic 2.628	NA	NA	NA
13353432	NA	TMhelix Region of a membrane-bound protein predicted to be embedded in the membrane, Phobius CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm, Phobius NON_CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region, Phobius TRANSMEMBRANE Region of a membrane-bound protein predicted to be embedded in the membrane.	NA	Inner Membrane 2.202	NA	NA	NA
13353433	NA	Phobius SIGNAL_PEPTIDE_N_REGION N-terminal region of a signal peptide, Phobius CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm, Phobius TRANSMEMBRANE Region of a membrane-bound protein predicted to be embedded in the membrane, TMhelix Region of a membrane-bound protein predicted to be embedded in the membrane.	NA	Outer Membrane 2.351	2gufA	0.91	89.08
13353434	NA	Phobius SIGNAL_PEPTIDE_N_REGION N-terminal region of a signal peptide, Twin-arginine translocation pathway, signal sequence, Phobius NON_CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region.	NA	Extracellular 1.888 Outer Membrane 1.815	2oy7A	7.4	86.59

13353435	NA	Phobius TRANSMEMBRANE Region of a membrane-bound protein predicted to be embedded in the membrane, TMhelix Region of a membrane-bound protein predicted to be embedded in the membrane, Phobius NON_CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the extracellular region, Phobius CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm.	NA	Inner Membrane 4.667	2nwlC	0.19	92.07
13353436	NA	Phobius TRANSMEMBRANE Region of a membrane-bound protein predicted to be embedded in the membrane, Phobius CYTOPLASMIC_DOMAIN Region of a membrane-bound protein predicted to be outside the membrane, in the cytoplasm, TMhelix Region of a membrane-bound protein predicted to be embedded in the membrane.	NA	Cytoplasmic 3.710	NA	NA	NA
13353437	ATP-binding cassette transporter nucleotide-binding domain; ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars, ions, peptides, and more complex organic molecules. The nucleotide-binding domain shows the highest similarity between all members of the family. ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region, in addition to, the Walker A motif/P-loop and Walker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins., HerA helicase [Replication, recombination, and repair]	AAA-like domain 505, P-loop containing nucleoside triphosphate hydrolase, SUPERFAMILY SSF52540	AAA-like domain	Cytoplasmic 2.972	2ipcA	0.094	70.25

Table 1: Characteristics of structure, function and sub-cellular localizations of the hypothetical proteins in Haloarchaeon Natrinema Sp. J7-2 plasmid.

We have successfully characterized probable functions of gene products by using CDD-Blast, Interproscan, and Pfam which was found to be 20, 28 and 11 respectively. Out of 48 screened hypothetical proteins, a total twenty-nine 3D structure prediction templates were also successfully characterized. The hypothetical proteins of NCBI gene ID 13353381 and 13353407 having arsenical resistance operon Repressor and metal regulated homodimeric repressors.

The bacterial transcription regulatory proteins (winged helix topology) from ARSR subfamily of helix-turn-helix include several proteins. These proteins dissociated from DNA in the presence of metal ions. The NCBI gene ID 13353399 of hypothetical protein shares a structurally conserved phosphate binding motif from TIM barrel proteins. This Conserved phosphate binding motif has eight beta/alpha closed barrel structure. The zinc-finger protein (ZPR1) is ubiquitous among eukaryotes and which is known to be an essential protein in yeast. ZPR1 is localized to the cytoplasm of quiescent cells. ZPR1 interacts with the cytoplasmic domain of the inactive Epidermal Growth Factor Receptor (EGFR). It inhibits the basal protein tyrosine kinase activity of EGFR (Table 1) [14].

Conclusion

In these studies total 48 functionally and structurally important hypothetical proteins from *Natrinema sp.* J7-2 plasmid has sorted. It is observed that many probable functional proteins are available in the *Natrinema sp.* J7-2 plasmid A total of 63 NCBI genes were screened proteins out which 48 hypothetical proteins which are successfully characterized by structurally as well as functionally using CDD- Blast, Interproscan, Pfam, Cello and PS2 server (available online web servers). The characterized predicted functions and three-dimensional structures in the bacterium life cycle can be assisting in establishing their role. The structural and functional characteristic of hypothetical proteins in Natrinema Sp. J7-2 plasmid was revealed from the present *in-silico* study. The data from this study may be useful for understanding the genetics and metabolic pathways at the molecular level in ecologically important haloarchaeon.

Acknowledgments

The first author wants to thanks, Swapnil Sanmukh, a postgraduate researcher at Sau Paulo state university, Brazil, for his indebted support and help during the study. Mr. Dilip Mesahram contributed in reviewing during the manuscript preparation. The authors are thankful to Scientist and Head, Water Technology and Management Division, CSIR-NEERI, Nehru Marg, Nagpur, for continuous support and valuable guidance during the study.

Bibliography

1. Feng Jie., *et al*. "The Complete Genome Sequence of Natrinema Sp. J7-2, a Haloarchaeon Capable of Growth on Synthetic Media without Amino Acid Supplements". *PloS One* 7.7 (2012): e41621.
2. Xin H., *et al*. "Natrinema Versiforme Sp. Nov., an Extremely Halophilic Archaeon from Aibi Salt Lake, Xinjiang, China". *International Journal of Systematic and Evolutionary Microbiology* 50.3 (2000): 1297-1303.
3. Lee Han-Seung. "Diversity of Halophilic Archaea in Fermented Foods and Human Intestines and Their Application". *Journal of Microbiology and Biotechnology* 23.12 (2013): 1645-1653.
4. Chen Shaoxing., *et al*. "Comparative Genomic Analysis of Halophiles Reveals New Clues to Their Adaptation Strategies in Hypersaline Environments". *Austin Journal of Proteomics, Bioinformatics and Genomics* 1.1 (2014).
5. Lv Jie., *et al*. "Isolation and Molecular Identification of Auxotrophic Mutants to Develop a Genetic Manipulation System for the Haloarchaeon Natrinema Sp . J7-2". *Archaea* (2015): 483194.
6. Zhang Ziqian., *et al*. "Temperate Membrane-Containing Halophilic Archaeal Virus SNJ1 Has a Circular dsDNA Genome Identical to that of Plasmid pHH205". *Virology* 434.2 (2012): 233-241.
7. Schäffer Alejandro A., *et al*. "Improving the Accuracy of PSI-BLAST Protein Database Searches with Composition-Based Statistics and Other Refinements". *Nucleic Acids Research* 29.14 (2001): 2994-3005.
8. Altschul Sf., *et al*. "Gapped BLAST and PSI- BLAST: A New Generation of Protein Database Search Programs". *Nucleic Acids Research* 25.17 (1997): 3389-3402.
9. Marchler-Bauer Aron., *et al*. "CDD: A Conserved Domain Database for Interactive Domain Family Analysis". *Nucleic Acids Research* 35.1 (2007): D237-D240.

10. Notredame C., *et al*. "T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment". *Journal of Molecular Biology* 302.1 (2000): 205-217.

11. Zdobnov EM and R Apweiler. "InterProScan--an Integration Platform for the Signature-Recognition Methods in InterPro". *Bioinformatics (Oxford, England)* 17.9 (2001): 847-848.

12. Bateman Alex., *et al*. "The Pfam Protein Families Database". *Nucleic Acids Research* 32 (2004): D138-D141.

13. Aydin Zafer., *et al*. "Protein Secondary Structure Prediction for a Single-Sequence Using Hidden Semi-Markov Models". *BMC Bioinformatics* 7 (2006): 178.

14. Siddiqui Azeem., *et al*. "Genome Annotation and Structure Predictions for Hypothetical Proteins in Agrobacterium Fabrum Str. C58 Plasmid At". *International Journal of Computer Applications* 85.1 (2014): 22-24.

Volume 3 Issue 1 November 2019
© All rights reserved by Hitesh S Thakare., *et al.*