IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 2, February 2022

DOI: 10.17148/IARJSET.2022.9276

Annotation of a hypothetical protein in space exposed *Bacillus pumilus SAFR_032* bacteria as Thioredoxin and structural analysis of the mutations

Neelambari D[1], Kusuma V K[2], Darshini R[3], Girish B[4], Rashmi kammar M[5], Manohar G M[6]

[1-6]Department of PG studies in Biotechnology, Nrupatunga University, Bangalore, Karnataka, India

[6]halliLabs, Ragihalli, Anekal Tq, Bangalore , Karnataka

Abstract : *Bacillus pumilus* strain *SAFR_032* is a spore forming bacteria that survived and showed high resistance when exposed to real space conditions (International space station). The space conditions included heat, ultraviolet, gamma, and cosmic radiations, desiccation and microgravity. Though the genome of space-exposed *B.pumilus SAFR_032* was sequenced some genes like BPUM_0788 and others were designated as hypothetical proteins . Using BLAST and domain databases we annotated BPUM_0788 gene as Thioredoxin gene, which is responsible for regulation of thioredoxin (TRX family) protein having oxidoreductase activity, peroxide resistance and oxidative stress response. These functions are known to help overcome stressful conditions. Comparison of BPUM_0788 with its counterpart from *Bacillus pumilus ATCC 7061* ground control in EMBOSS needle - pairwise alignment revealed three mutations specific to the space-exposed strain. These mutations are Alanine 83 serine (A83S), Glutamic acid 129 aspartic acid(E129D) and Aspartic acid 145 asparagine(D145N). Alpha fold 2 was used to model the structure of the protein. The functional and structural implications of these mutations were analyzed using domain search, STRING database and DynaMut2. Dynamut2 showed a destabilizing effect of all the three mutations. We hypothesize that these mutations may be helpful in thioredoxin to cope up with the stressful conditions experienced in space exposure. These methods can be used to annotate and analyse many remaining hypothetical proteins in the same and other bacteria.

Keywords: *Bacillus pumulis SAFR_032*, Hypothetical protein, Thioredoxin, gene annotation, mutation, space exposed stress

1.INTRODUCTION

Bacillus pumilus is a spore forming, gram positive, aerobic, rod-shaped, endospore forming bacterium. *Bacillus pumilus SAFR_032* which was isolated from the Jet Propulsion Laboratory–spacecraft assembly facility (JPL-SAF) showed high resistance and survived unfavorable space conditions such as ultrahigh vacuum (with its concomitant extreme desiccation), high and low extremes of temperature, solar UV radiation, cosmic radiation, and microgravity.[1][3]

An earlier study had found that during space exposure a set of proteins were reported to be upregulated in *Bacillus pumilus SAFR_032*. They are BL00422, alpha-ketoglutarate decarboxylase, 30s ribosomal protein S1, BPUM_0788, serine hydroxymethyltransferase, superoxide dismutase and glyceraldehyde-3-phosphate dehydrogenase.[1][3]. BL00422 and BPUM_0788 were designated as hypothetical proteins. In this study we annotated the BPUM_0788 hypothetical protein and studied mutations specific to the *B. pumilus SAFR 032* strain at the structural level. BLASTp search against the NCBI database identified BPUM-0788 as Thioredoxin protein. Thioredoxin has several family proteins, namely, Thioredoxin like protein, thioredoxin 1, 2, 3, reductase 1 & 3, perodoxin, etc.. Thioredoxin has oxidoreductase activity, peroxide resistance and oxidative stress response which are all needed most in stressful conditions. [4][6]

We studied the interaction of proteins between *Bacillus pumilus safr_032* and other proteins through a STRING database that imports data from experimentally derived protein–protein interactions through literature curation.[9]

Within a sequence, amino acids that are important for folding, structural stability, or that form a binding site are highly conserved and it is important to know the conservation status of residues to know their impact on the protein structure. A mutation in the conserved residues is more likely to change the protein structure and function. Conserved residues determine the evolutionary significance of the mutations. Using CONSURF (psiBlast and HMMER) we studied the conserved residues. Mutation of amino acid of *Bacillus pumilus* wild type and space strain showed conserved residues of Alanine to serine (A83S) ; we speculate that these residues might have altered the function of the protein.[11][12][13]

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 2, February 2022

DOI: 10.17148/IARJSET.2022.9276

Alpha fold 2 was used to build the structure of thioredoxin(BPUM_0788) protein in *Bacillus pumilus SAFR_032* and *Bacillus pumilus ATCC 7061* since the protein structures were not available in PDB.[14]
DynaMut2 is a web server used to analyze the effect of single amino acid changes (which can readily disrupt the structure and intermolecular interaction), stability, dynamics and even protein function and multiple mutations in protein structures. Dynamut2 analysis was done for the three mutations found in BPUM_0788 and their interactions with neighboring amino acids were charted out. The three mutations reduced the stability of the protein and showed different interactions with neighbouring amino acids.[15]
Our study is an investigation into the structural and functional role of annotated protein BPUM_0788 which was expressed during space conditions. We have tried to provide evidence for the role of mutations of this protein for increased resistance and survival of *B.pumilus SAFR_032.*

2. MATERIALS AND METHODS

2.1. Sequence retrieval : Protein sequence for the hypothetical protein (ABV61472.1) was retrieved from NCBI GenBank database. ABV61472.1 was used as a query in BLASTp (non redundant protein database, expect thresholds: 0.05) searched against non redundant database (https://blast.ncbi.nlm.nih.gov/Blast.cgi).[5]

2.2. Retrieval of interacting genes
Gene interactions of BPUM_0788 with other genes were retrieved from the STRING database. This database shows known interactions from curated databases, experimentally determined interactions, predicted interactions, gene neighborhood, gene fusions, gene co-occurrence and others.[9]

2.3. CONSURF TOOL
The conserved residues in the hypothetical protein BPUM_0788 from the space-exposed *Bacillus pumilus* strain were obtained by CONSURF(https://consurf.tau.ac.il/) using PsiBLAST and HMMER channels. Psi blast uses position-specific scoring matrices (PSSMs) to score matches between query and database sequences in subsequent iterations and HMMER searches sequence databases for sequence homologs, using probabilistic models called profile hidden Markov models (profile HMMs)[11][12][13]

2.4. ALPHA FOLD 2
The FASTA sequence of both BPUM_0788 (SAFR_032) and BAT_3297 (ATCC 7061) proteins were modeled using Alpha Fold 2. Alpha Fold 2 uses machine learning algorithms and artificial intelligence to model structures of unknown proteins constrained by multiple sequence alignments of homologs. The structures of these proteins were not available in the PDB-database.(https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipyn) [14]

2.5. DYNAMUT 2
The mutations of the hypothetical protein BPUM_0788 - A83S , E129D ,D145N were modeled in DynaMut2(http://biosig.unimelb.edu.au/dynamut). Hydrophobic interactions, clashes, van der-waals interactions and ionic interactions were analyzed using PyMol2. The interactions of the mutations and the distance of affected residues were measured in PyMol2.[15]

3. RESULTS :

Blast-p result showed that hypothetical protein BPUM_0788 which was expressed in *Bacillus pumilus-safr032* under space-exposed conditions is 100% identical with Thioredoxin protein. In the blast results we found 97.8% similarity between *Bacillus pumilus safr_032* and *Bacillus pumilus sp. 7788*. (Uniprot blastp).

The STRING database was used to retrieve protein-protein interaction between BPUM_0788 and other proteins.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 2, February 2022

DOI: 10.17148/IARJSET.2022.9276

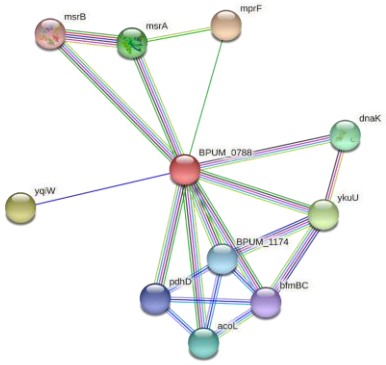


Fig. 2: BPUM_0778 interaction with other proteins form STRING database

Table. 1. List of protein names interacting with BPUM_0788

Sl.no.	Protein-protein interaction	Interacting protein names
1.	mprF	phosphatidylglycerol lysyltransferase
2.	yqiW	UPF0403 protein
3.	ykuU	putative peroxiredoxin
4.	msrA	peptide methionine sulfoxide reductase
5.	dnaK	chaperon protein
6.	acoL	dihydrolipoyl dehydrogenase
7.	BPUM_1174	dihydrolipoyl dehydrogenase
8.	pdhD	-
9.	bfmBC	-
10.	msrB	peptide methionine sulfoxide reductase

3.1 CONSURF

The conserved residues in the hypothetical protein BPUM_0788 from the space-exposed *Bacillus pumilus* strain were obtained by CONSURF using PsiBLAST and HMMER channels, which gave the following results (Table.2)

Table.2. CONSURF analysis showing the conservation status of mutations.

STRAINS	MUTATION AND POSITION	BURIED/ EXPOSED	CONSURF-PSI BLAST	CONSURF-HMMER BLAST
Bacillus pumilus ATCC7061	**ALANINE 83**	**BURIED**	**CONSERVED**	**AVERAGE**
	GLUTAMIC ACID129	EXPOSED	AVERAGE	VARIABLE
	ASPARTIC ACID145	EXPOSED	AVERAGE	VARIABLE
Bacillus pumilus	**SERINE 83**	**BURIED**	**CONSERVED**	**AVERAGE**

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 2, February 2022

DOI: 10.17148/IARJSET.2022.9276

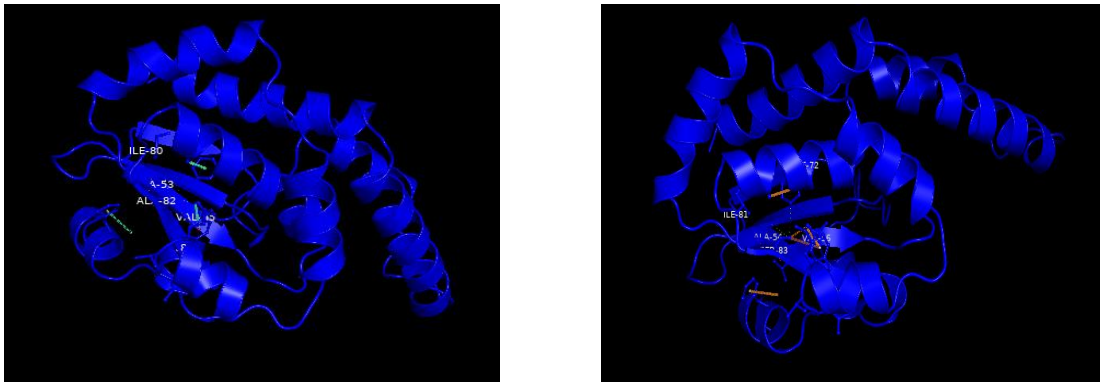
STRAINS	MUTATION AND POSITION	BURIED/ EXPOSED	CONSURF-PSI BLAST	CONSURF-HMMER BLAST
Bacillus pumilus ATCC7061	**ALANINE 83**	**BURIED**	**CONSERVED**	**AVERAGE**
	GLUTAMIC ACID129	EXPOSED	AVERAGE	VARIABLE
	ASPARTIC ACID145	EXPOSED	AVERAGE	VARIABLE
safr_032	ASPARTIC ACID 129	EXPOSED	AVERAGE	VARIABLE
	ASPARAGINE 145	EXPOSED	AVERAGE	VARIABLE

3.2 DynaMut 2

We used Dynamut2 to find out whether these mutations were causing any structural changes in the protein. Dynamut2 works on the basis of variation in the Gibbs free energy. and shows the prevalent bonds between amino acids in the protein. In our comparative analysis study between *Bacillus pumilus* wild type and space exposed strain, all the 3 mutations (A83S, E129D, D149N) that were unique to the space exposed strain were found to be of destabilizing in nature (Table 3.). The number of hydrophobic bonds also increased. Mutation from Alanine to Serine (A83S) makes protein -1.12 kcal/mol destabilizing, Glutamic acid to Aspartic acid (E129D) makes protein -0.6 kcal/mol destabilizing, Aspartic acid to Asparagine (D145N) makes protein -0.32 kcal/mol destabilizing. All the changes in free energies because of the three mutations indicate that these mutations destabilize the protein as a whole.. Even though the position at location 83 is buried in the proteins, the amino acidAlanine has mutated into serine.[6]

Table.3. Showing three mutations in the protein (BPUM_0788) and changes in energies accompanying them

MUTATION	CHANGE IN ENERGY
A83S	-1.12 kcal/mol
E129D	-0.6 kcal/mol
D145N	-0.32 kcal/mol



(a) (b)

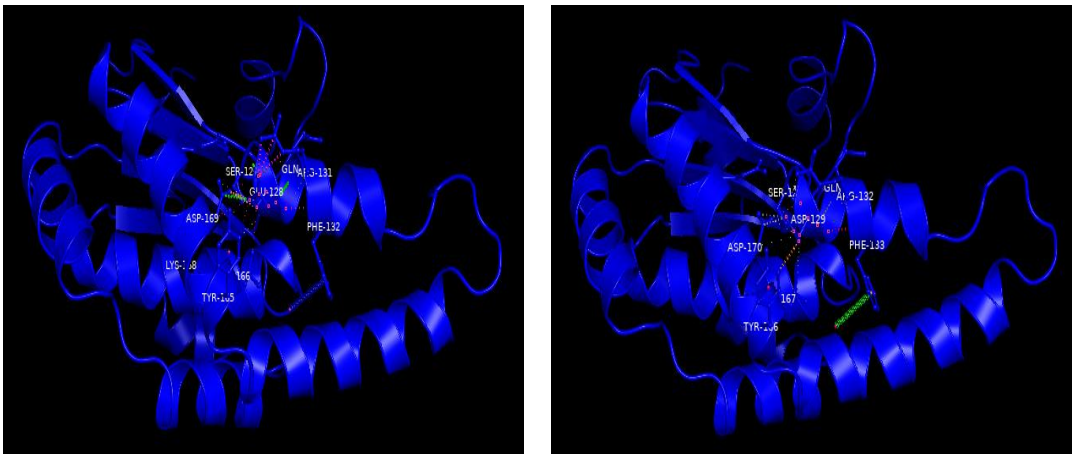
Fig.3. Interaction of (a)alanine-wild type and (b)serine-mutant type with other neighboring amino acids

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

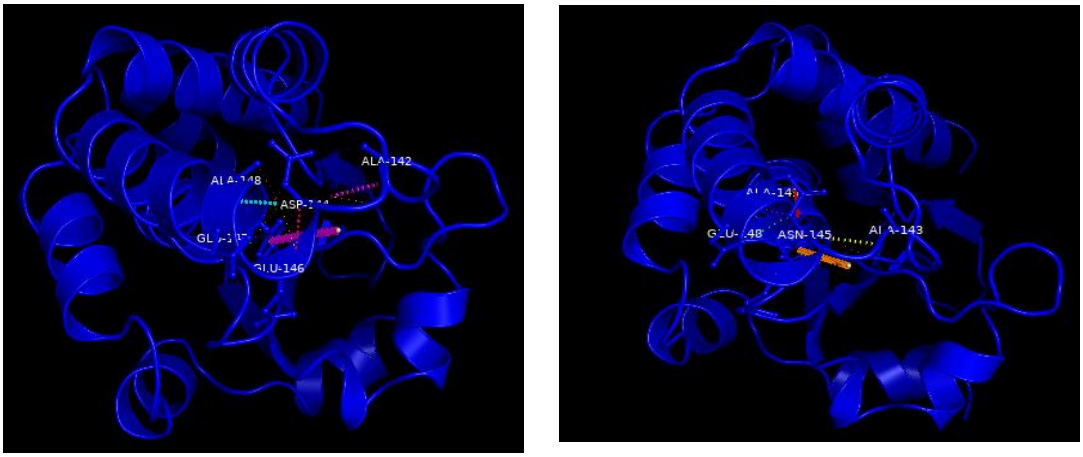
Impact Factor 7.105 ⁜ Vol. 9, Issue 2, February 2022

DOI: 10.17148/IARJSET.2022.9276



(c) (d)

Fig.4. Interaction of (c)glutamic acid-wild type and (d)aspartic acid-mutant type with other neighboring amino acids



(e) (f)

Fig .5. Interaction of (e)aspartic acid-wild type and (f)asparagine-mutant type with other neighboring amino acids

4. DISCUSSION :

We have annotated the hypothetical protein BPUM_0788 which was expressed in *Bacillus pumilus SAFR-032* when exposed to extreme conditions of space, as Thioredoxin protein. Along with BPUM_0788, other hypothetical proteins like BL00422 were also highly expressed on exposure to space conditions.. This thioredoxin has an oxidoreductase activity; it functions as thiol disulfide reductase that catalyze the reduction of protein disulfide bonds using an active site dithiol , present in CXXC motif .

We identified three mutations through comparative analysis between wild strain *Bacillus pumilus ATCC 7061* and space returned *Bacillus pumilus SAFR-032*. We hypothesized that these mutations might have helped the bacteria to survive in space conditions.

From the dynamut 2 server we analyzed that these mutations were destabilizing in nature. Though these mutations were destabilizing for some reasons they were tolerated in the protein itself suggesting that they may have been retained because of a hypothetical role in the function of Thioredoxin protein needed to tolerate space-exposed conditions. Thioredoxin has oxidoreductase activity, peroxide resistance and oxidative stress response which are all needed most in stressful conditions. Dynamut 2 analysis showed that there is an increase in the ionic interactions in the neighborhood . This might protect the protein structure during desiccation conditions. Further analysis such as distance calculation between the mutations, molecular interactions, and exact changes in the interaction at atomic level and function of hypothetical protein are required and may give the best results to understand the interactions which help the bacteria to survive in stressful conditions.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 7.105 ⁜ Vol. 9, Issue 2, February 2022

DOI: 10.17148/IARJSET.2022.9276

5. ACKNOWLEDGEMENTS:

We deeply appreciate the help and mentoring from **Dr. Naseer Pasha, PhD, NCBS** and Senior Scientist , Avesthagen, Bangalore, for regular inputs for this research.

6. REFERENCES :

[1]. Vaishampayan PA, Rabbow E, Horneck G, Venkateswaran KJ. Survival of Bacillus pumilus spores for a prolonged period of time in real space conditions. Astrobiology. 2012 May;12(5):487-97. doi: 10.1089/ast.2011.0738. PMID: 22680694.

[2]. Chiang AJ, Malli Mohan GB, Singh NK, Vaishampayan PA, Kalkum M, Venkateswaran K. Alteration of Proteomes in First-Generation Cultures of Bacillus pumilus Spores Exposed to Outer Space. mSystems. 2019 Jun 11;4(4):e00195-19. doi: 10.1128/mSystems.00195-19. PMID: 31186338; PMCID: PMC6561321.

[3]. Milojevic T, Weckwerth W. Molecular Mechanisms of Microbial Survivability in Outer Space: A Systems Biology Approach. Front Microbiol. 2020 May 15;11:923. doi: 10.3389/fmicb.2020.00923. PMID: 32499769; PMCID: PMC7242639.

[4]. Stepanov VG, Tirumalai MR, Montazari S, Checinska A, Venkateswaran K, Fox GE (2016) Bacillus pumilus SAFR-032 Genome Revisited: Sequence Update and Re-Annotation. PLoS ONE 11(6): e0157331. doi:10.1371/journal.pone.015733

[5]. Sayers EW, Karsch-Mizrachi I. Using GenBank. Methods Mol Biol. 2016;1374:1-22. doi: 10.1007/978-1-4939-3167-5_1. PMID: 26519398.

[6]. Zeller T, Klug G. Thioredoxins in bacteria: functions in oxidative stress response and regulation of thioredoxin genes. Naturwissenschaften. 2006 Jun;93(6):259-66. doi: 10.1007/s00114-006-0106-1.PMID: 16555095.

[7]. Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y, Raytselis Y, Sayers EW, Tao T, Ye J, Zaretskaya I. BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 2013 Jul;41(Web Server issue):W29-33. doi: 10.1093/nar/gkt282. Epub 2013 Apr 22. PMID: 23609542; PMCID: PMC3692093.

[8]. Ijaq, J., Malik, G., Kumar, A. et al. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. BMC Bioinformatics 20, 14 (2019).

[9]. Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, Marc Legeay, Tao Fang, Peer Bork, Lars J Jensen, Christian von Mering, Correction to 'The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets', Nucleic Acids Research, Volume 49, Issue 18, 11 October 2021, doi:10.1093/nar/gkab835

[10]. Schultz, L W et al. "The CXXC motif: crystal structure of an active-site variant of Escherichia coli thioredoxin." Acta crystallographica. Section D, Biological crystallography vol. 55,Pt 9 (1999): 1533-8. doi:10.1107/s907444999008756

[11]. Ben Chorin, Adi et al. "ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins." Protein science : a publication of the Protein Society vol. 29,1 (2020): 258-267. doi:10.1002/pro.3779

[12]. Altschul, S F et al. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic acids research vol. 25,17 (1997): 3389-402. doi:10.1093/nar/25.17.3389.

[13]. Finn, Robert D et al. "HMMER web server: interactive sequence similarity searching." Nucleic acids research vol. 39,Web Server issue (2011): W29-37. doi:10.1093/nar/gkr367.

[14]. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2. Epub 2021 Jul 15. PMID: 34265844; PMCID: PMC8371605.

[15]. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res. 2018 Jul 2;46(W1):W350-W355. doi: 10.1093/nar/gky300. PMID: 29718330; PMCID: PMC6031064.

Supplementary material :

1. BLAST
2. CONSURF
3. Alpha fold 2
4. DynaMut 2
5. Pairwise alignment