EUROPEAN JOURNAL OF BIOMEDICAL AND PHARMACEUTICAL SCIENCES
http://www.ejbps.com

ISSN 2349-8870
Volume: 3
Issue: 11
381-386
Year: 2016

CAPTURING THE STRUCTURAL VARIATION OF 1-PHOSPHOFRUCTOKINASE FROM DIFFERENT PATHOGENIC BACTERIA: AN *IN SILICO* APPROACH

Dr. Ayon Pal*

Department of Botany, Raiganj University, Raiganj – 733134, West Bengal, India.

***Author for Correspondence: Dr. Ayon Pal**
Department of Botany, Raiganj University, Raiganj – 733134, West Bengal, India.

Article Received on 07/09/2016 **Article Revised on 28/09/2016** **Article Accepted on 18/10/2016**

ABSTRACT
The enzyme phosphofructokinase (PFK) plays a vital role in the linear EMP glucose degradation pathway where it catalyzes an important rate determining step, the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate. In this study, a relative analysis of the PFK protein sequences and tertiary structures have been carried out from a few plant and human pathogenic eubacteria to capture how the amino acid compositional fluctuation has transpired into structural variation. Phylogenetic analysis of the PFK protein sequences was also carried out using Bayesian approach. Analysis of the protein sequence as well as the structural features of PFK from different pathogenic eubacterial species suggest that the enzyme has evolved in a polyphletic manner and has undergone variable evolutionary change. In terms of evolutionary change, the PFK enzyme of *Staphylococcus aureus* and *Streptococcus pneumoniae* was found to demonstrate the least amount of change. The PFK enzyme of *Clostridium perfringens* was found to be quite unique, both in terms of amino acid composition and structural features from the remaining eubacterial species considered in this study.

KEYWORDS: Phosphofructokinase, protein modeling, RMSD, multiple sequence alignment, phylogenetics, Bayesian inference, structural alignment, pathogenic bacteria.

1. INTRODUCTION

In the establishment of pathogenesis, glucose metabolism plays a substantial role and the metabolic needs of several pathogens are met by up regulation of host cell glycolysis. Down regulation of cellular glycolysis has also been reported in some circumstances.[1] Glycolysis is a linear sugar oxidation pathway involving ten stepwise chemical transformations through which glucose is metabolized into pyruvate. At the third step of glycolysis, key checkpoint exists, controlled by the enzyme phosphofructokinase (PFK). PFKs are tetrameric enzymes that have a key role in the regulation of glycolysis; as such, they are subject to allosteric activation and inhibition by various metabolites.[2] The PFK-catalyzed transfer of a phosphoryl group from ATP is an important reaction in a wide variety of biological processes.[3] Regulation of glycolytic enzymes is also known to occur by different mechanisms such as transcription control, mRNA stability modulation or by allosteric activation. One of the key regulating enzymes of the glycolytic pathway is phosphofructokinase (PFK), which is regulated by several kinases that are up regulated during infection.[4] PFK plays a vital role in the linear EMP glucose degradation pathway where it catalyzes an important rate determining step, the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate.[5] PFK subtypes are known to utilize either ATP, ADP, or pyrophosphate as the primary phosphoryl donor.[6] Although most enzymes of the glycolytic pathway are particularly conserved between different organisms, various types of phosphofructokinases exist with a very complex evolutionary history.[7] Crystal structure of PFK from the enteric bacteria *Escherichia coli* is found to comprise of two similar sections or lobes, where the alpha lobe is involved in ATP binding, and the beta lobe accommodates both the allosteric as well as the substrate-binding site.[8] The PFK enzyme being an integral part of a primeval but universal pathway like glycolysis[9], may be regarded as an excellent marker in determining the proximity among the different bacterial species at the metabolic level.

In this study a relative analysis of the PFK protein sequences and tertiary structures have been carried out from a few plant and human pathogenic eubacteria. The organisms considered in this study includes plant pathogenic *Pseudomonas syringae* and *Xanthomonas oryzae*. The human pathogenic organisms include *Acinetobacter baumanii, Staphylococcus aureus, Streptococcus pneumoniae, Mycoplasma pneumoniae, Clostridium perfringens* and *Vibrio cholera*. A detailed analysis of the composition profile of the PFK sequences

along with multiple protein sequence alignment was carried out to capture how the amino acid compositional fluctuation has transpired into sequence variation. Finally, an attempt has been made to find out how the sequence variation has moulded the three dimensional tertiary structure of the rate-limiting vital glycolytic enzyme PFK, and to what extent the structural variations are extended.

2. MATERIALS AND METHODS

The amino acid sequences of the eight selected bacterial PFK enzymes were obtained from GenBank hosted at http://www.ncbi.nlm.nih.gov/genbank.[10] The eight organisms from which the PFK amino acid sequences have been considered in this analysis is given in Table 1. These organisms represent both plant and human pathogenic eubacteria. Analysis of the amino composition of these PFK sequences were carried out using the ExPASy ProtParam tool.[11] The amino acid frequency of the PFK sequences were utilized for constructing a dendrogram based on UPGMA using DendroUPGMA with 100 bootstrap replicates (http://genomes.urv.cat/UPGMA).[12] Multiple sequence alignment of the eight PFK protein sequences was carried out using the multiple sequence alignment tool MUSCLE[13] hosted at http://www.ebi.ac.uk/Tools/msa/muscle.[14, 15] The multiple protein sequence alignment generated by MUSCLE was used as input for phylogenetic tree analysis using Bayesian methods. MrBayes[16] was utilized for tree inference involving Bayesian method. MrBayes performs Bayesian inference of phylogeny using a variant of Markov chain Monte Carlo.[16] The likelihood model for MrBayes analysis was set up with GTR substitution type, Blosum62 substitution model with Invariable+gamma rates variation across sites. The Markov Chain Monte Carlo parameters were set at 100000 generations with tree sampling frequency of 100. The phylogenetic tree were inferred using the online server Phylogeny.fr hosted at http://www.phylogeny.fr.[17] The three dimensional (3D) structure prediction of the PFK protein enzyme and analysis was carried out using the Phyre2 suite of tools available at http://www.sbg.bio.ic.ac.uk/phyre2.[18] Phyre2 is a protein homology/analogy recognition engine which uses advanced remote homology detection methods to build 3D models, predict ligand binding sites and analyze the effect of amino acid variants for a user's protein sequence.[18] Multiple structural alignment of the predicted PFK protein 3D structure and RMSD calculation was carried out employing MISTRAL. RMSD analysis gives a good idea about the structural relatedness of the protein tertiary structures. MISTRAL implements a strategy for multiple protein alignment based on the minimization of an energy function over the low-dimensional space of the relative rotations and translations of the molecules.[19]

3. RESULTS AND DISCUSSION

The enzyme PFK has been found to display tremendous amino acid compositional variability[20], and fluctuations in the relative frequency of the different amino acids is certain to introduce variations in sequence composition. In this analysis, the amino acid sequences of PFK enzyme from eight different plant as well as human pathogenic eubacteria have been thoroughly studied in terms of their amino acid composition pattern, and the data obtained is presented in Table 2. It is clearly evident from Table 2, that there is a considerable difference in the frequency of the twenty standard amino acids constituting the PFK enzyme in different bacterial species. A UPGMA based dendrogram utilizing the relative frequency of the twenty amino acids comprising the PFK sequences from the eight pathogenic bacteria was constructed and shown in Figure 1. This diagram shows that from the compositional perspective, PFK sequences from the wall-less *Mycoplasma* is quite distinct from the rest of the other organisms. Another interesting observation is the clubbing of the PFK sequence of the gram negative *Acinetobacter baumanii* with the clade bearing the PFK sequences of the gram positive *Streptococcus*, *Staphylococcus* and *Clostridium*.

In order to have an idea about how the amino acid frequency fluctuation has shaped up into sequence variability, a multiple protein sequence alignment was carried out. A multiple protein sequence alignment of the eight sequences using MUSCLE[13] was constructed and phylogeny inference using Bayesian method was carried out. This was done to garner idea about the relation between the different PFK sequences and also to study their evolutionary divergence. The phylogram constructed using MrBayes given in Figure 2, and it shows a polytomy at the base of the tree with three distinct speciation events. The PFK enzyme of *Streptococcus pneumoniae* and *Clostridium perfringens* is found to evolve separately along with a third clade that contains the PFKs of *Staphylococcus aureus*, *M. pneumoniae*, *X. oryzae*, *V. cholerae*, *A. baumanii* and *P. syringae.*

The PFK enzyme sequence of *C. perfringens* was foumd to demonstrate the greatest evolutionary change, and standout from the rest of the other PFK protein sequences both in terms of amino acid frequency as well as sequence variability. *Mycoplasma pneumoniae* was also found to display high degree of evolutionary change in its PFK enzyme. In terms of evolutionary change, the PFK enzyme of *Staphylococcus aureus* and *Streptococcus pneumoniae* demonstrates the least amount of change. The PFK of the gram negative *P. syringae*, *A. baumanii*, *V. cholera* and *X. oryzae* was found to display somewhat similar degree of evolutionary change.

In order to find out how the PFK sequence variability has translated into structural variability, 3D protein modeling of the PFK sequences from the eight selected pathogenic

bacterial species was carried out using the Phyre2 web portal.[18] Model quality assessment of the eight PFK models was carried out and a multiple structural alignment of the obtained models was carried out using the web server MISTRAL.[19] The multiple structural alignment of the eight PFK tertiary structures captured an alignment core size of 86 residues and the mean square distance of the aligned residues was 2.0 (computed over 2408 corresponding pairs of 86 residues in 8 proteins). A pairwise structural alignment details of the eight PFK tertiary structures is given in Table 3. From Table 3, it is evident that the PFK enzyme of *C. perfringens* is structurally quite different from the rest of the other bacterial PFKs considered in this study. The RMSD calculation between the pairwise PFK tertiary structure shows that the PFK of *C. perfringens* has a greater RMSD deviation (2.3Å to 2.7Å) with all the concerned PFK tertiary structures included in this study. A diagrammatic representation of the PFK tertiary structure from *C. perfringens* and *P. syringae* is given in Figure 3. The PFK tertiary structure of *A. baumanii* and *P. syringae* was found to have a low RMSD score of 0.9Å which is in line with their phylogenetic relatedness. The RMSD between the PFK tertiary structures of *P. syringae* and *V. cholera* was calculated to be the lowest (0.3Å) among all the PFK tertiary structures suggesting strong structural similarity between the two. The alignment energy between the two structures was also found to be the lowest (-1115.381) among all the pairwise structure comparisons thus, corroborating the similarity between the two structures at the three dimensional level. The PFK tertiary structures of *Staphylococcus aureus* and *Streptococcus pneumoniae* was also found to display a larger RMSD deviation of 2.2Å. This finding supports the phylogenetic inference, where the PFK sequences were found to evolve in three separate lineages, with *Staphylococcus aureus* and *Streptococcus pneumoniae* as residents of two separate clades.

Table 1: List of the organisms whose PFK protein sequences have been analyzed in this study.

Organism	Family	Pathogenic to	Gram Nature
Pseudomonas syringae	Pseudomonadaceae	Plants	Negative
Xanthomonas oryzae	Xanthomonadaceae	Plants	Negative
Acinetobacter baumanii	Moraxellaceae	Human beings	Negative
Staphylococcus aureus	Staphylococcaceae	Human beings	Positive
Streptococcus pneumoniae	Streptococcaceae	Human beings	Positive
Mycoplasma pneumoniae	Mycoplasmataceae	Human beings	Wall less
Clostridium perfringens	Clostridiaceae	Human beings	Positive
Vibrio cholerae	Vibrionaceae	Human beings	Negative

Table 2: Relative amino acid usage frequency of the PFK amino acid sequences from the eight organisms selected for this study.

Amino acid	*Pseudomonas syringae*	*Xanthomonas oryzae*	*Acinetobacter baumanii*	*Staphylococcus aureus*	*Streptococcus pneumoniae*	*Mycoplasma pneumoniae*	*Clostridium perfringens*	*Vibrio cholerae*
Ala	12.78	15.72	8.89	9.48	6.97	7.33	9.72	10.36
Cys	0.00	1.26	0.64	0.98	0.91	0.33	0.63	1.79
Asp	4.47	6.60	4.76	6.86	5.46	6.33	5.96	4.64
Glu	5.75	3.77	6.98	5.56	6.06	2.67	8.15	5.71
Phe	2.88	1.26	3.49	4.25	4.24	4.00	2.19	3.21
Gly	9.59	9.12	6.35	8.17	6.36	4.67	11.29	8.21
His	2.56	2.52	3.49	0.65	2.73	1.00	1.57	1.79
Ile	4.47	3.46	8.57	9.80	9.70	4.33	9.09	4.64
Lys	3.83	0.94	5.71	6.54	6.06	6.00	6.27	4.64
Leu	13.10	13.21	10.16	6.21	9.39	15.67	8.78	11.79
Met	1.92	1.89	1.59	0.98	1.82	2.67	3.45	2.50
Asn	2.24	3.15	5.08	4.90	6.36	7.00	3.45	4.29
Pro	4.15	5.03	3.81	3.27	3.94	4.33	1.57	2.86
Gln	6.07	4.72	6.98	4.58	5.46	8.67	1.25	7.86
Arg	4.15	7.55	2.54	1.63	2.12	2.00	5.33	3.57
Ser	5.43	5.35	5.40	6.86	7.58	3.33	5.33	5.71
Thr	6.39	5.35	6.35	8.17	6.36	6.00	6.27	5.36
Val	9.27	7.55	7.62	9.15	5.46	9.33	7.84	8.57
Trp	0.96	0.94	0.64	0.00	0.30	0.67	0.00	2.14
Tyr	0.00	0.63	0.95	1.96	2.73	3.67	1.88	0.36

Table 3: Details of the multiple structural alignment of the eight PFK tertiary structures modeled using Phyre2. The multiple PFK protein structural alignment was carried out using MISTRAL.

Structure 1	Structure 2	Str 1_ Length	Str 2_ length	Alignment Energy	No. of matches	RMSD (Å)	Sequence Identity	z-score	p-value
A. baumanii	*C. perfringens*	305	318	-304.061	69	2.7	7	1.2	1.17E-01
A. baumanii	*M. pneumoniae*	305	293	-1015.281	278	1.2	63	11.3	2.83E-07
A. baumanii	*P. syringae*	305	307	-1102.471	304	0.9	161	12.4	6.56E-08
A. baumanii	*S. aureus*	305	306	-768.927	233	2	66	7.8	2.51E-05
A. baumanii	*S. pneumoniae*	305	316	-933.959	278	1.5	77	9.8	2.03E-06
A. baumanii	*V. cholerae*	305	310	-1104.881	304	0.8	131	12.3	7.42E-08
A. baumanii	*X. oryzae*	305	308	-909.2	274	1.4	89	9.7	2.20E-06
C. perfringens	*M. pneumoniae*	318	293	-308.776	81	2.6	7	1.2	1.08E-01
C. perfringens	*P. syringae*	318	307	-318.485	79	2.5	5	1.4	9.20E-02
C. perfringens	*S. aureus*	318	306	-294.092	65	2.7	11	1	1.37E-01
C. perfringens	*S. pneumoniae*	318	316	-280.799	100	2.5	9	0.9	1.69E-01
C. perfringens	*V. cholerae*	318	310	-329.861	90	2.3	9	1.5	7.61E-02
C. perfringens	*X. oryzae*	318	308	-307.794	81	2.5	11	1.2	1.10E-01
M. pneumoniae	*P. syringae*	293	307	-1026.342	284	1.4	69	11.4	2.57E-07
M. pneumoniae	*S. aureus*	293	306	-701.033	221	2	43	6.9	8.51E-05
M. pneumoniae	*S. pneumoniae*	293	316	-894.048	266	1.6	54	9.2	4.07E-06
M. pneumoniae	*V. cholerae*	293	310	-1032.807	285	1.4	67	11.4	2.67E-07
M. pneumoniae	*X. oryzae*	293	308	-879.85	261	1.5	58	9.3	3.72E-06
P. syringae	*S. aureus*	307	306	-776.664	238	2.1	60	7.9	2.27E-05
P. syringae	*S. pneumoniae*	307	316	-933.668	280	1.5	66	9.8	2.04E-06
P. syringae	*V. cholerae*	307	310	-1115.381	306	0.3	142	12.5	6.16E-08
P. syringae	*X. oryzae*	307	308	-916.357	277	1.4	93	9.8	1.94E-06
S. aureus	*S. pneumoniae*	306	316	-833.606	264	2.2	84	8.4	1.17E-05
S. aureus	*V. cholerae*	306	310	-775.632	232	2	62	7.8	2.58E-05
S. aureus	*X. oryzae*	306	308	-824.042	264	2.1	72	8.5	1.01E-05
S. pneumoniae	*V. cholerae*	316	310	-935.461	278	1.5	74	9.8	1.97E-06
S. pneumoniae	*X. oryzae*	316	308	-1088.914	301	2.1	61	11.9	1.35E-07
V. cholerae	*X. oryzae*	310	308	-919.194	276	1.5	94	9.8	2.01E-06

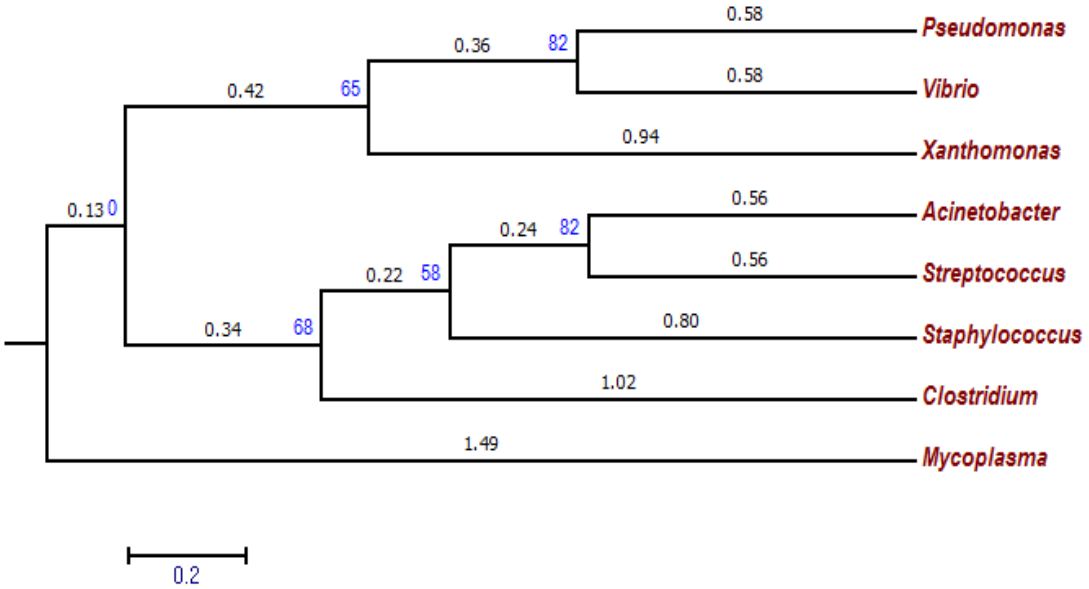

Figure 1: A dendrogram constructed utilizing the relative frequency of the twenty standard amino acids from the PFK sequences of the eight pathogenic bacterial species selected for this study.

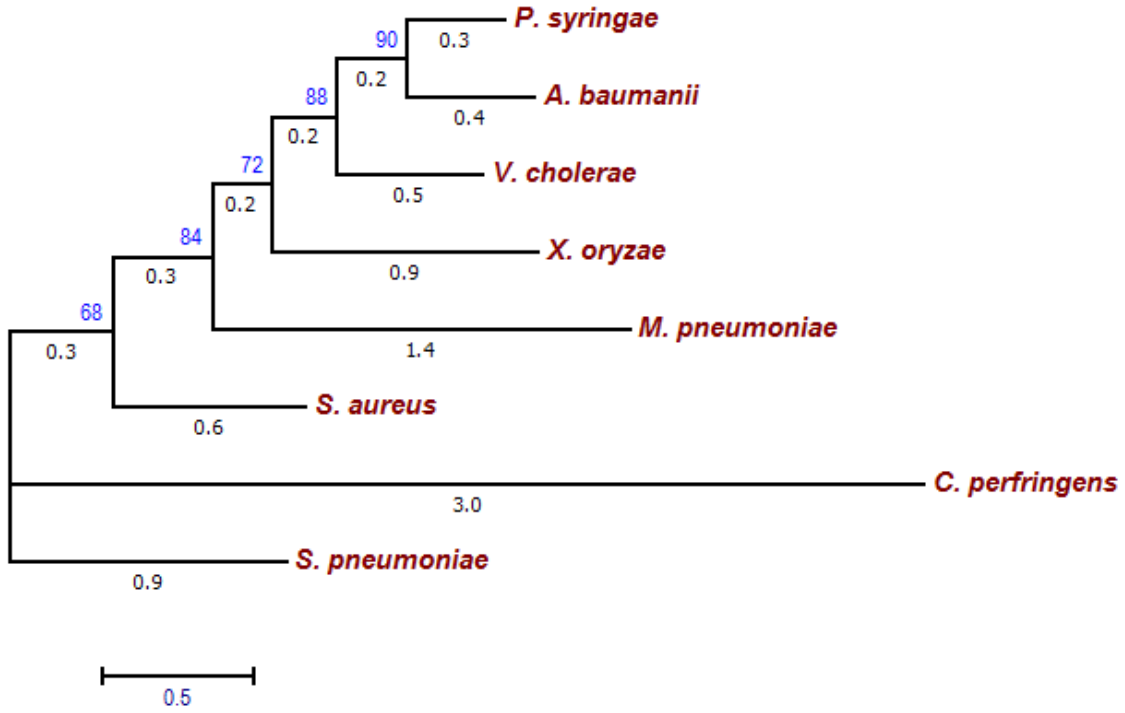

Figure 2: An unrooted phylogenetic tree obtained from the PFK protein sequences of eight pathogenic eubacteria after multiple sequence alignment with MUSCLE. This tree is generated using MrBayes.

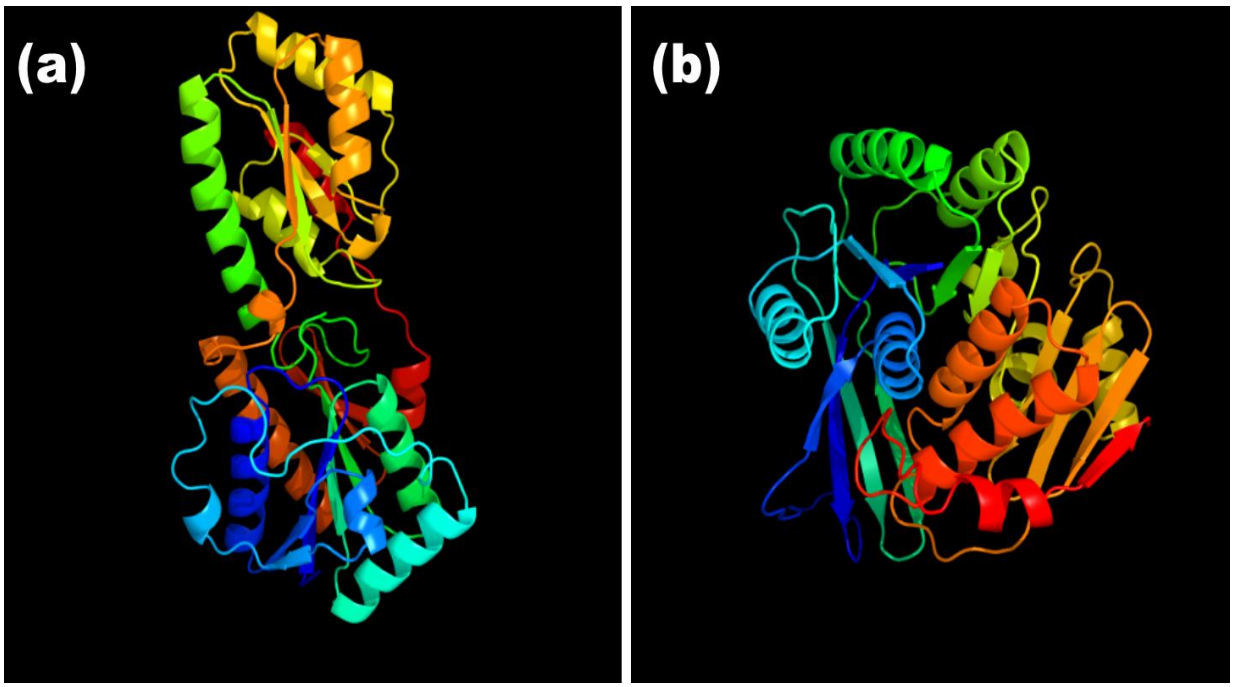

Figure 3: Ribbon depiction of the 3D tertiary structures of PFK modeled from (a) *Clostridium perfringens* and (b) *Pseudomonas syringae* using Phyre2.

4. CONCLUSIONS

The enzyme phosphofructokinase (PFK) is a vital component of the glycolytic pathway. The results from an earlier as well this study points to the fact that, PFK is subjected to quite a greater degree of composition bias irrespective of the taxonomic status. Analysis of the protein sequence as well as the structural features of PFK from different pathogenic eubacterial species suggest that the enzyme has evolved in a polyphyletic manner and has undergone variable evolutionary change. In terms of evolutionary change, the PFK enzyme of *Staphylococcus aureus* and *Streptococcus pneumoniae* was found to demonstrate the least amount of change. As suggested from this analysis, the PFK of *Clostridium perfringens* was found to be quite unique, both in terms of amino acid composition and structural features, from the rest of the organisms considered in this study.

5. REFERENCES

1. Rojas, C.M., et al., Regulation of primary plant metabolism during plant-pathogen interactions and its contribution to plant defense. Frontiers in Plant Science, 2014; 5: 17.
2. Poorman, R.A., et al., Evolution of phosphofructokinase--gene duplication and creation of new effector sites. Nature, 1984; 309(5967): 467-9.
3. Hellinga, H.W. and P.R. Evans, Mutations in the active site of Escherichia coli phosphofructokinase. Nature, 1987; 327(6121): 437-9.
4. Ghukasyan, V.V. and A.A. Heikal, Natural Biomarkers for Cellular Metabolism: Biology, Techniques, and Applications., 2014; CRC Press.
5. Furuya, E. and K. Uyeda, An activation factor of liver phosphofructokinase. Proc Natl Acad Sci U S A, 1980; 77(10): 5861-4.
6. Ronimus, R.S. and H.W. Morgan, The biochemical properties and phylogenies of phosphofructokinases from extremophiles. Extremophiles, 2001; 5(6): 357-73.
7. Bapteste, E., D. Moreira, and H. Philippe, Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase. Gene, 2003; 318: 185-91.
8. Shirakihara, Y. and P.R. Evans, Crystal structure of the complex of phosphofructokinase from Escherichia coli with its reaction products. J Mol Biol, 1988; 204(4): 973-94.
9. Pal, A., et al., Genomic heterogeneity within conservedmetabolic pathways of Arthrobacter species - a bioinformatic approach. Bioinformation, 2011; 5(10): 446-54.
10. Benson, D.A., et al., GenBank. Nucleic Acids Res, 2013; 41(Database issue): D36-42.
11. Walker, J.M., The Proteomics Protocols Handbook., 2005; Humana Press.
12. Garcia-Vallvé, S., A. Romeu, and J. Palau, Horizontal Gene Transfer in Bacterial and Archaeal Complete Genomes. Genome Research, 2000; 10(11): 1719-1725.
13. Edgar, R.C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research, 2004; 32(5): 1792-1797.
14. Li, W., et al., The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic acids research, 2015; 43(W1): W580-4.
15. McWilliam, H., et al., Analysis Tool Web Services from the EMBL-EBI. Nucleic acids research, 2013; 41(Web Server issue): W597-600.
16. Huelsenbeck, J.P. and F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics, 2001; 17(8): 754-5.
17. Dereeper, A., et al., Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res, 2008; 36(Web Server issue): W465-9.
18. Kelley, L.A., et al., The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protocols, 2015; 10(6): 845-858.
19. Micheletti, C. and H. Orland, MISTRAL: a tool for energy-based multiple structural alignment of proteins. Bioinformatics, 2009; 25(20): 2663-9.
20. Ayon, P., Compositional Variability Analysis of Phosphofructokinase Enzyme Primary Sequence Within the Eubacterial Domain. International Journal of Current Research in Life Sciences, 2016. Accepted for publication.