International Journal of Current Research in Life Sciences
Vol. 05, No. 03, pp. 554-561, March 2016

www.ijcrls.com

Full Length Research Article

COMPOSITIONAL VARIABILITY ANALYSIS OF PHOSPHOFRUCTOKINASE ENZYME PRIMARY SEQUENCE WITHIN THE EUBACTERIAL DOMAIN

***Ayon Pal**

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

Accepted 02nd February 2016; Published Online 31st March 2016

ABSTRACT

Glycolysis performs ten stepwise chemical transformations and at the third step of glycolysis, major checkpoint exists, regulated by the enzyme phosphofructokinase (PFK). The PFK enzyme being an integral part of a primeval but universal pathway may be regarded as an excellent marker in determining the proximity among the different bacterial species at the metabolic level. In this study an attempt was made to capture the extent of compositional variability that has been accommodated by the PFK enzyme sequence, yet keeping its primary functional attribute intact. In this study more than eight thousand PFK amino acid sequences belonging to 782 bacterial genera comprising the entire eubacterial domain were comprehensively analyzed. The variability in the amino acid usage pattern existing within the PFK enzyme was found to be astonishing, and resolutely reflected in the GRAVY as well as the pI values. Barring a few amino acids such as histidine, cysteine and tryptophan, which are in general less frequent, nearly all the other amino acids showed a significant fluctuation within the compositional profile of PFK. Even at the generic level substantial compositional variability was detected from this analysis. *Mycoplasma* and *Acinetobacter* species were found to be extremely opposite in nature considering their PFK compositional profile.

Key words: Phosphofructokinase, Glycolysis, Eubacteria, Amino acid usage, Isoelectric point, GRAVY, Compositional variability.

INTRODUCTION

Glycolysis is the first process in the cellular combustion of glucose which is the source of almost all energy used by cells and occurs in nearly all living organisms within the cytosol (Romano and Conway, 1996). Glycolysis is also important because the metabolism of glucose produces useful intermediates for other metabolic pathways, such as the synthesis of amino acids or fatty acids. Glycolysis performs ten stepwise chemical transformations and at the third step of glycolysis, major checkpoint exists regulated by the enzyme phosphofructokinase (PFK). The PFK-catalyzed transfer of a phosphoryl group from ATP is an important reaction in a wide variety of biological processes (Hellinga and Evans, 1987). 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 (Furuya and Uyeda, 1980). Whereas most glycolytic enzymes are remarkably conserved between different organisms, various types of phosphofructokinases exist with a very complex evolutionary history (Bapteste, Moreira, and Philippe, 2003). Crystal structure of PFK from *Escherichia coli* have been shown to be comprised of two similar sections or lobes. The alpha lobe is involved in ATP binding and the beta lobe houses both the substrate-binding and the allosteric site (Shirakihara and Evans, 1988).

Corresponding author: Ayon Pal,
Department of Botany, Raiganj University, Raiganj – 733134, West Bengal, India.

This study attempts to capture the degree of variability persisting within the amino acid sequence of the PFK enzyme at the amino acid compositional level. The PFK enzyme being an integral part of a primeval but universal pathway (Pal, Mondal, Mukhopadhyay, and Bothra, 2011), may be regarded as an excellent marker in determining the proximity among the different bacterial species at the metabolic level. Since, protein function is a virtue of its structure, which in turn is dictated by the amino acid composition, in this study an attempt was made to capture the extent of compositional variability that has been accommodated by the PFK enzyme sequence, yet keeping its primary functional attribute intact. An important facet of this elaborate study was to capture this compositional variability by studying more than eight thousand PFK amino acid sequences belonging to 782 bacterial genera comprising the entire eubacterial domain. These genera represent the different categories of bacteria known to man, both from the taxonomic as well as metabolic viewpoint.

MATERIALS AND METHODS

In order to conduct this elaborate study 8248 amino acid sequences of the PFK enzyme was obtained from GenBank (Benson *et al.,* 2013). The amino acid frequency of the twenty different amino acids constituting the PFK protein primary sequence in the 8248 sequences was computed using our own script developed in Python. The grand average of hydropathy (GRAVY) score was calculated using GRAVY CALCULATOR hosted at http://www.gravy-calculator.de and

the theoretical isoelectric point (pI) was calculated using the Compute pI tool hosted at http://web.expasy.org/compute_pi. The grand average of hydropathicity or GRAVY (Kyte and Doolittle, 1982) of the linear polypeptide sequence is calculated as the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence. Increasing positive score indicates greater hydrophobicity. The calculation is based on the Kyte-Doolittle scale (Kyte and Doolittle, 1982). It is a simple method for displaying the hydropathic character of a protein. Isoelectric point (pI) is a pH in which the net charge of protein is zero. In case of proteins, isoelectric point mostly depends on seven charged amino acids: glutamate (δ-carboxyl group), aspartate (ß-carboxyl group), cysteine (thiol group), tyrosine (phenol group), histidine (imidazole side chains), lysine (ε-ammonium group) and arginine (guanidinium group).The pI value can affect the solubility of a molecule at a given pH and can be used to distinguish proteins based on their compositional variability.

RESULTS AND DISCUSSION

Analysis of the PFK enzyme from 8248 sequences demonstrated the enzyme to be composed of an average about 314 amino acid residues with an average molecular weight of 33597.21 Da. The highest molecular weight was demonstrated by *Mycoplasma parvum* (36710 Da) followed by *Uneaplasma parvum* (36499 Da), *Rhodopirellula sallentina* (36338 Da), *Sulfuricurvum kujiense* (36237 Da), *Mycoplasma iowae* (36338 Da) and *Arcobacter anaerophilus* (36278 Da). On the other hand, species like *Actinosynnema minum* (29960 Da), *Amycolatopsis azurea* (30031 Da), *Streptomyces decoyicus* (30018 Da) and *Mycobacterium smegmatis* (30218 Da) demonstrated molecular weights at the lower end of the spectrum.

Amino acid composition profile of PFK

The usage frequency of the 20 different amino acid residues constituting the PFK enzyme sequence was thoroughly analyzed. Amino acid wise analysis of the 8248 PFK sequences showed the following characteristic features:

- **Alanine:** The relative frequency of alanine was found to range from 1.974% for *Mycoplasma iowae* to 24.149% for *Rubrivivax benzoatilyticus*. Elevated levels of alanine was also evident in the case of organisms like *Xanthomonas*, *Delftia*, *Belnapia*, *Novispirillum*, *Streptomyces spp.*, *Mycobacterium spp.*, *Arthrobacter spp.,* etc. Alanine is a non-polar, aliphatic amino acid which demonstrated the highest range of fluctuation in usage within the PFK sequences.
- **Cysteine:** The relative frequency of cysteine was found to range from as low as 0.303% in different species of *Agrococcus*, *Arthrobacter*, *Corynebacterium*, *Deinococcus* to 3.7% in species like *Eubacterium desmolans Enterorhabdus caecimuris*, *Ruminococcus sp.* and others. Cysteine was found to be absent in 263 species that included different species of *Microbacterium*, *Deinococcus*, *Pseudomonas*, *Halomonas*, *Thermus*, *Staphylococcus*, *Lactobacillus*, *Acinetobacter*, *Listeria*, *Clostridium* and other species like *Nocardia farcinica*, *Corynebacterium accolens*, *Sinomonas humi* and others.
- **Aspartic acid:** The relative frequency of this acidic amino acid was found to range between 1.94% and 10.10%. Different species of *Rufibacter* and *Bacillus* was found to have lower percentage of aspartic acid (~2%) compared to organisms like *Jiangella alkaliphila* and *Citreicella sp.* (~10%).

Table 1. Comparative details of the relative frequency of the twenty different amino acids constituting the PFK enzyme primary structure in 8248 sequences

Amino acids	Min	Freq. of minimum	Max	Freq. of maximum	1st Quartile	Median	3rd Quartile
Alanine	1.974	2	24.149	1	7.812	9.180	11.576
Cysteine	0.000	496	3.704	2	0.625	0.938	1.294
Aspartic acid	1.942	1	10.095	1	4.878	5.751	6.562
Glutamic acid	1.852	1	12.662	1	5.573	6.562	7.500
Phenylalanine	0.000	16	7.643	2	2.159	3.096	3.988
Glycine	3.514	7	15.409	1	8.882	9.901	11.315
Histidine	0.000	130	5.902	1	0.990	1.618	2.303
Isoleucine	0.000	1	14.375	1	4.531	7.261	9.091
Lysine	0.306	2	15.161	1	3.560	5.281	6.667
Leucine	3.727	1	17.143	1	7.927	9.615	11.075
Methionine	0.303	4	5.648	1	1.320	1.942	2.606
Asparagine	0.307	2	12.500	1	2.752	3.762	4.872
Proline	0.926	1	10.577	1	2.326	3.165	4.516
Glutamine	0.000	18	11.182	1	1.881	2.913	4.348
Arginine	0.321	1	12.539	1	2.951	4.918	6.250
Serine	1.582	2	14.985	1	4.934	5.714	6.604
Threonine	2.469	1	11.881	1	5.000	5.769	6.562
Valine	2.812	1	16.460	1	7.143	8.464	9.571
Tryptophan	0.000	2488	3.185	1	0.000	0.323	0.667
Tyrosine	0.000	344	6.230	2	1.238	2.174	2.821

Table 2. The synonymous codons of the five different amino acids whose abundance is positively correlated with alanine in the 8248 PFK sequences

Amino acid	Side chain polarity	Codons	Spearman correlation coefficient (ρ) at $p<0.01$
Phenylalanine	Non-polar	UUU, UUC	−0.56
Isoleucine	Non-polar	AUU, AUC, AUA	−0.67
Lysine	Polar	AAA, AAG	−0.79
Asparagine	Polar	AAU, AAC	−0.60
Tyrosine	Polar	UAC, UAU	−0.62

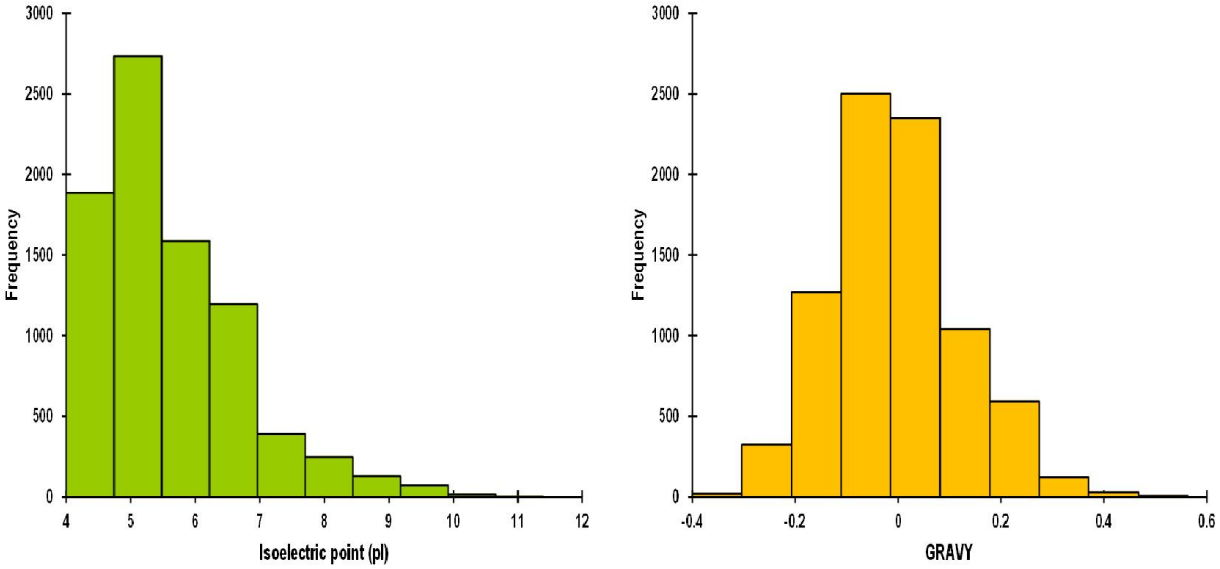

Figure 1. Histogram showing the frequency distribution of isoelectric point (pI) and GRAVY score of the 8248 PFK sequences from 2654 different species belonging to 782 bacterial genera

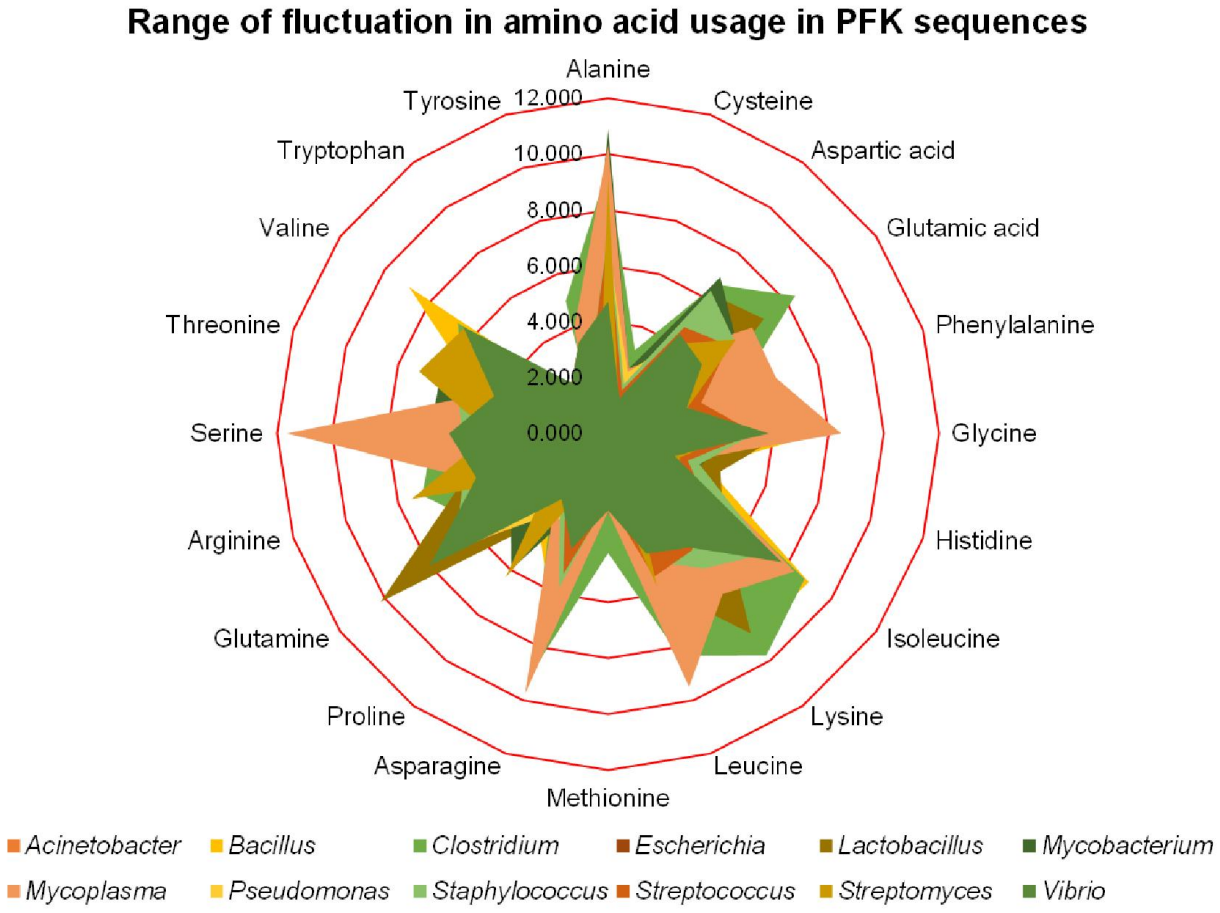

Figure 2. A 3D radar plot showing the comparative range of fluctuation in the usage of the twenty different amino acids within the amino acid sequences of the PFK enzyme from the selected twelve eubacterial genera. The figures over the concentric circles indicate the range in fluctuation of amino acid usage expressed in percentage

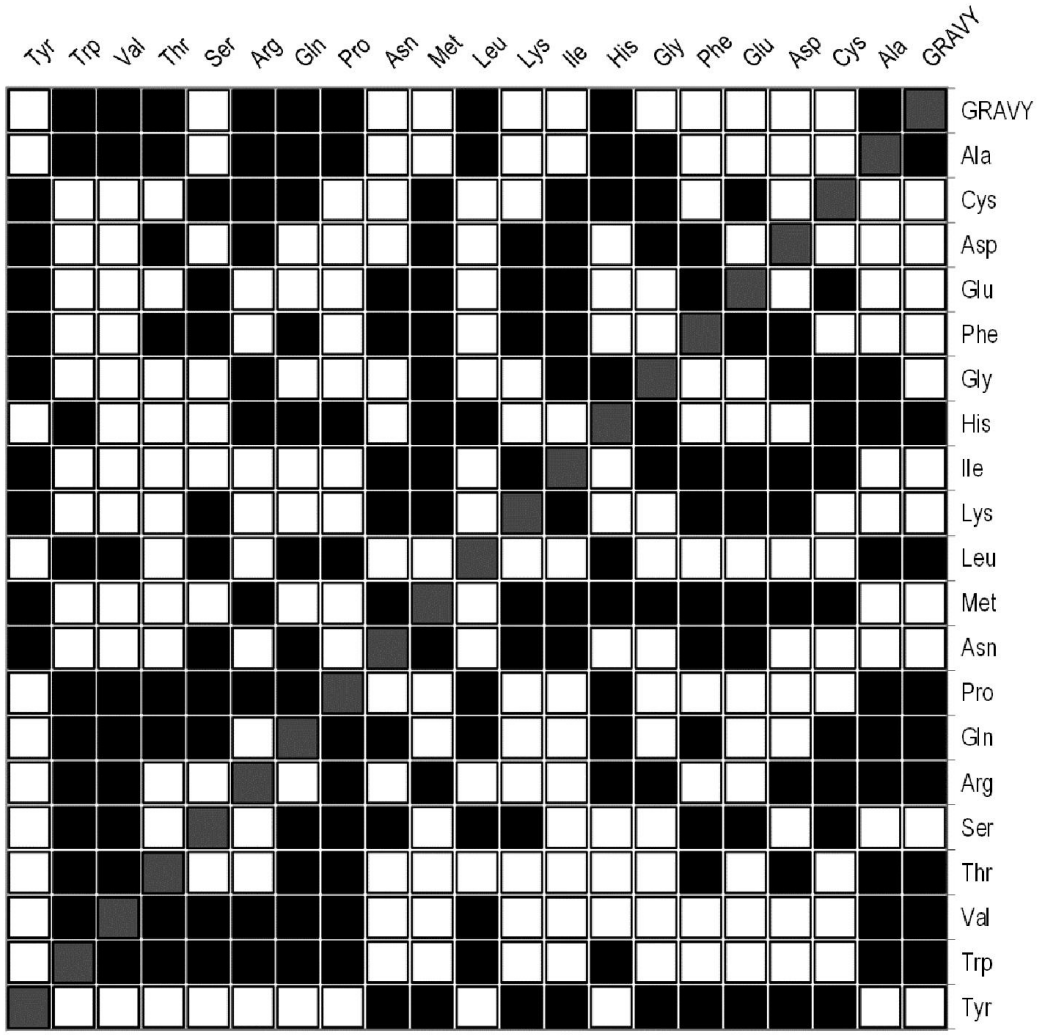

Figure 3. A correlation map showing the relation between the twenty different amino acids constituting the 8248 PFK sequences from 2654 different species belonging to 782 bacterial genera. Black and white squares indicates positive and negative correlations respectively whereas the diagonal (ρ=1, p<0.05) is represented in grey

- **Glutamic acid:** Similar to aspartic acid, the relative frequency of this polar negatively charged, aliphatic amino acid was also found to range from 1.8% to 12.67%. *Citreicella sp.* and *Streptomyces griseoflavus* demonstrated lower amounts of glutamic acid (1.8%), whereas *Clostridium cadaveris*, *Helococcus kunzii* and *Fusobacterium necrophorum* displayed higher amounts of glutamic acid (~12%).
- **Phenylalanine:** This neutral, non-polar amino acid was found to be absent in 16 species which included organisms like *Actinopolyspora erythraea*, *Nocardiopsis lucentensis*, *Actinomyces slackii*, different species of *Gordonia*, *Mycobacterium*, *Nocardia* and others. The relative frequency of phenylalanine in PFK was found to range from 0.3% in *Mycoplasma hyorhinis* to about 7.64% in different species of *Fusobacterium*.
- **Glycine:** This relative frequency of this amino acid was found to be the lowest in the PFK sequences of the different species of *Mycoplasma* (about 3.5%) like *M. mycoides*, *M. leachii*, *M. capricolum*, *M. hyorhinis*, *M. iowae* and *Clostridium spp*. Compositionally, the relative frequency of glycine was greatest in organisms like *Rhodococcus sp.* (15.4%), *Deinococcus gobiensis* (14.9%) and *Bacillus flexus* (14.7%).
- **Histidine:** This positively charged, aromatic amino acid was found to be absent in the PFK sequences of atleast 60 different species that included *Aurantiomonas manganoxydans*, *Fusobacterium ulcerans*, *F. mortiferum*, *F. varium*, *Staphylococcus haemolyticus*, *Clostridium acetobutylicum*, *Borrelia burgdorferi*, *Staphylococcus caprae*, *Mycoplasma hyorhinis* and others.
- **Isoleucine:** The relative frequency of this non-polar, uncharged, aliphatic amino acid was found to be the lowest in different species of *Nocardiopsis* (0.32%). Organisms like *Buchnera aphidicola* and *Peptoniphilus* depicted highest levels of isoleucine in their PFK sequences. *Actinosynnema mirum* was the only exception in this category since it lacked isoleucine in its PFK.
- **Lysine:** The lowest frequency of lysine was observed in the organisms like *Cellulosimicrobium cellulans* (0.30%) followed by *Actinosynnema mirum*, *Agromyces subbeticus* and others. Higher levels of lysine (~15%) was found to be present in the PFK of *Halanaerobium praevalens* and *Dictyoglomus turgidum.*
- **Leucine:** Organisms like *Butyrivibrio fibrisolvens* (3.7%), *Belliella baltica* (4.3%) demonstrated the lowest amount of leucine compared to different species of *Chromobacterium*

(17.14%), *Azotobacter* and *Mycoplasma* that have a higher frequency of this nonpolar amino acid in their PFK.

- **Methionine:** This S-methyl thioether side chain containing non-polar, aliphatic amino acid had the lowest usage frequency (0.30%) in the PFK of different species of *Arthrobacter*, *Rhodococcus*, *Cellulomonas* and others. The frequency of methionine was highest in the case of organisms like *Oribacterium* (5.65%), *Blautia* (5.3%), *Mesotoga* (5.01%), etc.
- **Asparagine:** The relative frequency of this polar, aliphatic amino acid was found to range from 0.3% in different species of *Agromyces* and *Pseudomonas stutzeri* to 12.5% in the different species of *Mycoplasma* including *M. iowae*.
- **Proline:** Different species of *Clostridium* like *C. arbusti*, *C. pasteurianum*, etc. demonstrated low proline content (about 0.9%) whereas, higher amounts of proline in PFK was observed in many species of *Streptomyces* like *S. xiaopingdaonensis* (10.6%) and different species of *Mycobacterium*.
- **Glutamine:** Out of the 8248 different PFK amino acid sequences, eleven different species belonging to the genera *Brachyspira*, *Clostridium*, *Fusobacterium* and *Burkholderia pseudomallei* were found to lack glutamine. Lower levels of glutamine (about 3%) was demonstrated by *Rhodobacter sphaeroides*, *Aquimarina atlantica*, *A. agarilytica*, *Cellulomonas gilvus*, *Azospirillium spp.* and others. Higher levels of glutamine (about 11%) was detected in the PFK sequences of *Lactobacillus mellifer*, *Acinetobacter haemolyticus*, *A. baumanii*, *Vibrio cholera*, *Mycoplasma pneumonia*, *Pseudomonas fuscovaginae*.
- **Arginine:** The frequency of this charged, aliphatic amino acid in the PFK sequences was found to vary from 0.32% demonstrated by *Spiroplasma diminutum* and *Entomoplasma somnilux* to 12.54% as observed in *Streptomyces griseoflavus* and different species of *Burkholderia*.
- **Serine –** PFK of *Belnapia moabensis* was found to contain the lowest amount of serine, a polar amino acid (1.6%) in comparison to different species of *Mycoplasma* such as *M. suis*, *M. parvum* and *M. wenyonii*. An interesting observation was the elevated levels of serine (about 10%) in a large number of *Bacillus* species such as *B. coagulans*, *B. aquimaris* and *B. vietnamensis*.
- **Threonine:** *Alicyclobacillus acidocaldarius*, along with *Chromobacterium subtsugae, Brevibacillus borstelensis*, *Clostridium butyricum*, *C. bifermentans* and *Pseudomonas aeruginosa* depicted lowest levels of threonine residues (about 2.5%) in their PFK. On the other hand, highest level of threonine was observed in *Leuconostoc lactis* (11.88%) followed by *Rhodococcus sp.*, *Corynebacterium durum*, *C. kutscheri* and many species of *Staphylococcus*.
- **Valine:** Lowest level of valine was found in the PFK of *Eubacterium saphenum* (2.8%) whereas *Rhodococcus sp.* (16.46%) and *Mobilicoccus spp.* demonstrated higher levels of valine incorporated in their PFK.
- **Tryptophan:** The frequency of this heaviest amino acid in the composition of PFK was very low (0.30%) in case of organisms like *Mycobacterium thermoresistible*, *Yonghaparkia sp.*, and *Eubacterium cellulosolvens*. *Alicyclobacillus acidocaldarius* demonstrated the highest frequency of tryptophan (3.185%) followed by organisms like *Morganella morani*, *Escherichia coli*, *Salmonella enterica*, *Gilliamella apicola*, *Providencia spp*. 2488 PFK sequences belonging to 1231 different species was found to lack tryptophan. Some of these organisms include different species of *Bacillus* like *B. cereus, B. weihenstephanensis, B. thuringiensis, B. flexus, B. megaterium* and organisms like *Staphylococcus warneri, S. pasteuri, Lactobacillus melliventris* and many others.
- **Tyrosine:** 344 PFK sequences belonging to 108 different species was found to lack tyrosine residues. Some of these organisms include *Chloroflexus aggregans, Pleomorphomonas oryzae, P. koreensis* and different species of *Rubrivivax, Xanthomonas, Actinoplanes, Chromobacterium* and others. Very low levels of tyrosine in the PFK was observed in case of *Vibrio tubiashii* (0.30%). *Clostridium botulinum* demonstrated the highest level (6.23%) of tyrosine in their PFK sequences.

A comparative account of the relative frequencies of the twenty standard amino acids constituting the 8248 PFK enzyme sequences of the 782 bacterial genera is given in Table 1. From Table 1 it is quite evident that amino acids such as cysteine, histidine, tyrosine and particularly tryptophan are not indispensable in the PFK sequences of many organisms. The lower quartile score of tryptophan frequency was found to be zero which is significantly different from the other low usage amino acids. The relative frequency of alanine was found to encompass a larger range (1.974% to 24.15%) which is highest among all the amino acids. Alanine cannot be phosphorylated, has a linear confirmation and the methyl group of alanine is non-reactive. Because of this non-reactivity, alanine is almost never directly involved in protein function and hence its higher compositional fluctuation within the PFK sequences is supposed to alter the enzymes functionality at the minimum. In contrast, the fluctuation in relative distribution of tryptophan is the lowest (0.0 to 3.185%) followed by cysteine (0.0 to 3.704%).

Comparative GRAVY and pI analysis of PFK

A comparative study of the GRAVY score along with the pI was carried out taking into account the 8248 PFK sequences from the 782 bacterial genera. Both these parameters can be used to distinguish between proteins based on their physico-chemical attributes. The GRAVY index indicates the solubility of the proteins, where the positive GRAVY indicates hydrophobic nature whereas the negative GRAVY reflects hydrophilic property of the protein. The GRAVY score of the PFK sequences was found to range between -0.39 and 0.55, exemplified by *Rhodopirellula sallentina* and *Rhodobacter capsulatus* respectively. A histogram showing the frequency distribution of GRAVY scores from 8248 PFK sequences is given in Figure 1. The highest number of sequences were found to be within the GRAVY range of −0.11 to −0.01 (2501 sequences) and −0.14 to −0.08 (2350 sequences). Thus, in terms of solubility, a majority 54.67% of the PFK sequences (4510 sequences) were found to be hydrophilic. 39 PFK sequences belonging to 26 different species were found to display a GRAVY score of zero. These included organisms like *Geobacillus caldoxylosilyticus, Planomicrobium glaciei, Pasteurella dagmatis, Kluyvera ascorbata, Vibrio parahaemolyticus, Bacillus cereus, Streptococcus parasanguinis, Vibrio vulnificus, Chryseobacterium sp., Peptoclostridium difficile, Vibrio tasmaniensis, Borrelia burgdorferi, Desulfosporosinus*

meridiei, Escherichia coli, Pedobacter arcticus, Pseudomonas aeruginosa, Shigella dysenteriae, Shigella sonnei, Pedobacter kyungheensis, Hymenobacter norwichensis, Shimwellia blattae, Collinsella intestinalis, Weissella cibaria, Listeria newyorkensis and *Klebsiella pneumonia*.

Isoelectric point (pI) is a pH in which net charge of a protein is zero. The pI of the PFK sequences was found to range between 4.02 and 11.35, exemplified by *Actinomyces slackii* and *Streptomyces griseoflavus* respectively. The pI value thus show a staggering range of fluctuation. Figure 1 shows that a large number of PFK sequences display a pI value which is within the acidic pH range of 4.74 and 5.48. A meagre 622 PFK sequences were found to have pI within the basic pH range of 7.2 and 11.35. About 190 PFK sequences belonging to 160 different species demonstrated pI values in the neutral pH range of 7 to 7.2. These included species like *Actinobacillus equuli, Bacillus alveayuensis, Vibrio anguillarum, Vibrio cholera, Vibrio navarrensis, Vibrio ordalii, Vibrio vulnificus, Proteus vulgaris, Haemophilus paraphrohaemolyticus, Clostridium pasteurianum, Mycoplasma ovipneumoniae, Bacillus selenitireducens, Clostridium saccharogumia, Clostridium botulinum, Borrelia burgdorferi, Proteiniclasticum ruminis, Psychromonas arctica, Xenorhabdus doucetiae* and others.

Compositional variability of PFK at the generic level

In order to better understand the compositional variability of the PFK enzyme primary structure, a critical analysis was carried out considering some selected bacterial genera such as *Acinetobacter, Bacillus, Clostridium, Escherichia, Lactobacillus, Mycobacterium, Mycoplasma, Pseudomonas, Staphylococcus, Streptococcus, Streptomyces* and *Vibrio*. This was done to gain an insight into the degree of variability existing in the relative amino acid residual composition of PFK at the generic level.

- ***Acinetobacter*:** The amino acid composition data of 118 PFK sequences from *Acinetobacter* demonstrated the absence of cysteine in 12 sequences. The range of distribution of tryptophan was found to be the lowest (0.335%) whereas, it was highest for lysine (4.14%) followed by glutamine (3.68%) and isoleucine (3.51%). The relative frequency of leucine usage (8.86% to 11.93%) was found to be the highest among all the amino acids.
- ***Bacillus:*** Analyzing the 437 PFK sequences from different species of *Bacillus* it was observed that about 58% (254) of the sequences were lacking tryptophan. The range of alanine usage frequency was the highest for alanine (9.80%) followed closely by isoleucine, lysine, valine and leucine (all about 8.50%). The relative frequency of glycine usage (7.05% to 14.73%) was found to be the highest among all the amino acids.
- ***Clostridium:*** Out of the 328 PFK sequences from different *Clostridium spp.* cysteine was found to absent in 24 sequences. Tryptophan was found to be absent in 69% of the *Clostridium* PFK sequences (226 sequences). The range of dispersion in lysine usage (9.76%) was found to be the highest followed by alanine (9.54%). The relative frequency of glycine and isoleucine usage (about 5% to 14.0%) was found to be the highest among all the amino acids.
- ***Escherichia:*** Analysis of the 196 PFK sequences from *Escherichia* did not show the absence of any particular amino acid. In comparison to the other three aforementioned genera, the range of usage of the twenty amino acids was found to be quite narrower in case of *Escherichia*. In contrast to alanine and leucine, the range of usage was found to be higher for isoleucine (6.128%) followed by glycine (5.28%) and glutamine (5.22%). The relative frequency of alanine, leucine and glycine usage (about 7% to 12.0%) was found to be the highest among all the amino acids.
- ***Lactobacillus*:** 259 PFK sequences from *Lactobacillus* were analyzed and 93 of these did not contain tryptophan. Cysteine was also found to be missing in 49 of the PFK sequences from *Lactobacillus.* The highest dispersion in case of usage was displayed by alanine (10.63%) followed closely by glutamine (10.20%) and lysine (8.80%). Low range of usage was detected in case of tryptophan (1.29%) and cysteine (2%). The relative frequency of leucine and glycine usage (about 7% to 13.0%) was found to be the highest among all the amino acids.
- ***Mycobacterium:*** Analyzing the 70 PFK sequences from *Mycobacterium*, the usage of alanine was found to be largely scattered (10.9%) followed by aspartic acid (6.9%), serine and threonine (6.5%). The relative usage of alanine (10.3% to 21.20%) and valine usage (9% to 12.7%) was found to be the highest among all the amino acids.
- ***Mycoplasma:*** Out of the 111 PFK sequences from *Mycoplasma* a few were found to lack cysteine, tryptophan and histidine. Of the twenty amino acids, the fluctuation in the usage of serine, alanine, asparagine and leucine were found to be highest (about 10.5%). The relative usage of lysine (6% to 13%) and leucine usage (6% to 15.7%) was found to be the highest among all the amino acids.
- ***Pseudomonas:*** Out of the 348 PFK sequences from *Pseudomonas*, tyrosine was found to be absent in 267 sequences. Cysteine was also found to be absent from 98 PFK sequences. Alanine and glutamine showed the highest fluctuation in usage (about 7%). The relative usage of alanine (10.8% to 17.5%) and leucine usage (11.4% to 15.8%) was found to be the highest among all the amino acids.
- ***Staphylococcus:*** 138 PFK sequences of *Staphylococcus* out of the 209 were found to lack tryptophan residues followed by cysteine in 65 sequences. Highest fluctuation in usage was found in the case of alanine (7.2%) and isoleucine (7.21%). The relative usage of glycine (7.09% to 12.42%) was found to be the highest among all the amino acids.
- ***Streptococcus*:** Analyzing 502 PFK sequences from *Streptococcus*, alanine and valine was found to show the highest fluctuation in usage (about 7%). Both glycine and leucine was found to be utilized in larger proportions about 7.5 to 12.5%.
- ***Streptomyces*:** In this genera, out of the 244 PFK sequences, the range of fluctuation was found to be highest in case of alanine (9.3%) and arginine (7.5%). The relative usage of alanine (12.16% to 21.40%) and glycine (9.7% to 14.51%) was found to be the highest among all the amino acids.
- ***Vibrio*:** Out of the 349 PFK sequences from the different *Vibrio* spp., tryptophan was found to be absent in 49 PFK sequences. The range of fluctuation in usage was found to

be higher for glutamine (8%) and isoleucine (7.77%). The relative usage of leucine (8.12% to 12.62%) and glycine (7.3% to 13.12%) was found to be the highest among all the amino acids.

A 3D radar plot showing the fluctuation in the usage of the twenty amino acids in the twelve selected bacterial genera is given in Figure 2. From this plot it may be assumed that *Mycoplasma* has a very distinct amino acid usage profile compared to the other eleven selected genera. The range of glycine and serine usage in the amino acid sequence of the PFK of from different species of *Mycoplasma* show a comparatively different pattern from the rest of the genera considered in this study. The usage pattern of serine in *Mycoplasma spp.* was found to vary up to 12% which is the highest for all the amino acids among the selected genera. The amino acid sequence of PFK from different species of *Vibrio* was found to demonstrate a compact range in the usage of the different amino acids which suggests the conserved nature of amino acid utilization within this genera. The amino acid composition profile of PFK of the different species of *Clostridium* was also found to vary quite extensively with respect to asparagine, lysine and alanine. The PFK sequences from the different species of *Bacillus* demonstrated the highest range of fluctuation in the usage of the branched chain, non-polar, hydrophobic amino acid valine.

The amino acid composition of PFK from different species of the acid-fast bacterium *Mycobacterium* demonstrated the highest fluctuation in the usage of alanine. Alanine is regarded as a hydrophobic ambivalent molecule, which means that it can be either inside or outside of the protein molecule. The amino acid profile of PFK from different species of *Acinetobacter* was found to display the most constricted or alternatively, the most conserved usage. The genus *Acinetobacter* represent strictly aerobic, non-fermentative, gram-negative bacilli. All the PFK sequences derived from the different species of this genera were found to display amino acid variation within the 4% level which is the lowest in comparison to the other selected species.

The most variable or less conserved amino acid utilization pattern was detected in the PFK sequences from the gram positive *Bacillus, Clostridium, Lactobacillus* and the cell wall less *Mycoplasma*. The gram positive *Staphylococcus* and *Streptomyces* was also found to display moderately higher variability in terms of amino acid usage. In general, the gram negative species were found to display lower levels of fluctuation in their PFK amino acid profile compared to the gram positive species.

Correlation analysis of relative amino acid usage in PFK sequences

To assess the association or correlation among the usage of the twenty different amino acids a Spearman rank order correlation analysis was carried out taking all the 8248 PFK sequences from the 2654 different species belonging to 782 bacterial genera comprising the eubacterial domain. Spearman correlation analysis considering the twenty amino acids and the GRAVY score (which is a physical property assessing parameter) at $p<0.01$ significance level revealed the strong positive association of GRAVY score with the non-polar amino acids alanine (ρ=0.72), leucine (ρ=0.52), proline (ρ=0.62) and valine (ρ=0.53). The compositional frequency of tyrosine, one of the universally low usage amino acid in the 8248 amino acid sequences of PFK, was found to be significantly positively correlated with isoleucine (ρ=0.69), leucine (ρ=0.60) and negatively correlated with alanine (ρ=−0.62), methionine (ρ=−0.58), proline (ρ=−0.60), valine (ρ=−0.56) and tryptophan (ρ=−0.61) which is the lowest utilized amino acid. The frequency of alanine residues in the PFK sequences, which is a non-polar, aliphatic and one of the highly utilized amino acid, was found to be largely anti-correlated with the non-polar amino acids such as phenylalanine (ρ=−0.56), Isoleucine (ρ=−0.67) and polar amino acids such as lysine (ρ=−0.79), asparagine (ρ=−0.60) and tyrosine (ρ=−0.62). A common feature of all these amino acids are that they are coded by two synonymous codons (except isoleucine which has three codons) and all these synonymous codons vary only at the third position (Table 2). A correlation map showing the relationship between the twenty amino acids constituting the 8248 PFK sequences from 2654 different species belonging to 782 bacterial genera along with the GRAVY score is given in Figure 3.

The variability in the amino acid usage pattern existing within the PFK enzyme is remarkable, which is resolutely reflected in the GRAVY as well as the pI values. Barring a few amino acids such as histidine, cysteine and tryptophan, which are in general less frequent, nearly all the other amino acids showed a significant fluctuation within the compositional profile of PFK. Even at the generic level substantial compositional variability was detected from this analysis. Apart from genera like *Acinetobacter, Pseudomonas* and *Escherichia*, most of the other genera considered in this study like *Mycoplasma, Bacillus, Clostridium, Lactobacillus, Staphylococcus* and *Vibrio* demonstrated significant intra-generic amino acid usage variability within their PFK primary structure. Amongst all the genera considered in this study, *Bacillus, Clostridium* and *Mycoplasma* standout from rest of the other genera because of their high fluctuation levels considering a majority of the amino acids. From the amino acids utilization perspective, glycine, alanine, valine, lysine, glutamic acid, lysine, aspartic acid and isoleucine show the maximum fluctuations in usage level within the PFK primary structure. The genus *Acinetobacter* appeared to be the most unique group of organism, where a very conservative amino acid utilization pattern was detected. The compositional variability within this genus was the lowest in comparison to the rest of the genera and apart from lysine no other amino acid was found to fluctuate beyond four percent level. The GRAVY and pI analysis also confirms the finding regarding the compositional variability and both the GRAVY score as well as the pI values were found to encompass a large range. All these findings suggest the PFK enzyme to be in a dynamic state of evolution which has also been detected to some degree in its genic properties (Bapteste *et al.,* 2003).

Conclusion

The PFK enzyme is a vital component of the glycolytic pathway and within the bacterial domain this enzyme has been subjected to a great deal of compositional variability. This study was carried out on 8248 phosphofructokinase enzyme sequences and it was observed that PFK has sustained a

greater degree of variability in its amino acid composition but these variations have been accommodated or absorbed in a strategic manner to cope up with the lifestyle of the organism, keeping the enzyme's function intact.

Acknowledgements

The author would like to acknowledge Prof. Subhasis Mukhopadhyay and Dr. Asim Kumar Bothra for their constructive suggestions and Mrs. Monalisha Pal for assistance in carrying out this study.

REFERENCES

Bapteste, E., Moreira, D., and Philippe, H. 2003. Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase. *Gene, 318*, 185-191.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. 2013. GenBank. *Nucleic Acids Res, 41*(Database issue), D36-42. doi: 10.1093/nar/gks1195

Furuya, E., and Uyeda, K. 1980. An activation factor of liver phosphofructokinase. *Proc Natl Acad Sci U S A, 77*(10), 5861-5864.

Hellinga, H. W., and Evans, P. R. 1987. Mutations in the active site of Escherichia coli phosphofructokinase. *Nature, 327*(6121), 437-439. doi: 10.1038/327437a0

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol, 157*(1), 105-132.

Pal, A., Mondal, U. K., Mukhopadhyay, S., and Bothra, A. K. 2011. Genomic heterogeneity within conservedmetabolic pathways of Arthrobacter species - a bioinformatic approach. *Bioinformation, 5*(10), 446-454.

Romano, A. H., and Conway, T. 1996. Evolution of carbohydrate metabolic pathways. *Res Microbiol, 147*(6-7), 448-455.

Shirakihara, Y., and Evans, P. R. 1988. Crystal structure of the complex of phosphofructokinase from Escherichia coli with its reaction products. *J Mol Biol, 204*(4), 973-994.

****