

Research Article

Homology-Based Functional Relationships of Uncharacterized G8 Protein of *Homo sapiens*

Padma Saxena

Department of Zoology, D.A.V. College, Civil Lines, Kanpur-208001, Uttar Pradesh, India

*Corresponding author

Padma Saxena

Email: padmasaxenadav@gmail.com

Abstract: The G8 (for eight conserved glycine residues) domain is widely distributed, being found in *Homo sapiens* is analyzed by sequence similarity search using Blast and Clustal W. It reveals uncharacterized homologues, found in *Pan troglodytes*, *Nomascus leucogenys*, *Pongo abelii* and, *Nomascus leucogenys*. Data base searching results indicate that G8 protein of *Homo sapiens* is closely related to *Pan troglodytes*, *Nomascus leucogenys*, *Pongo abelii* and, *Nomascus leucogenys*. Phylogenetic analysis reveals that G8 protein of human shows maximum similarity with four uncharacterized protein of *Pan troglodytes*, *Nomascus leucogenys*, *Pongo abelii* and, *Nomascus leucogenys*. Discovery of G8 domain and its homologues will be important for the research of the structure/function of related proteins and beneficial for the development of novel therapeutics.

Keywords: G8 domain, Data base searching, Blast, Clustal W, Protein sequence, Homology.

INTRODUCTION

Proteins are building block of living organism and involved in virtually all cell functions. Each protein performs specific function within the body. They are constructed from a set of 20 amino acids and have distinct three-dimensional shapes [1]. The G8 domain is widely distributed, being found in proteins from various animals (from *Strongylocentrotus purpuratus* to *Homo sapiens*), lower eukaryotes (such as *Dictyostelium discoideum* and *Tetrahymena thermophila*) and bacteria (such as the alpha-proteobacteria *Nitrobacter hamburgensis*, gamma-proteobacterium *Hahella chejuensis* and the green non-sulfur bacterium *Chloroflexus aurantiacus*) including some hereditary disease related protein such as PKHD1, KIAA1109 and TMEM2 proteins [2-4] but not present in plants, viruses and archaea. These proteins are integral membrane proteins with signal peptides and/or transmembrane segments, with two well-conserved glycine residues and the PbH1 domain [5]. They are members of the large family of ATP-binding cassette (ABC) transporters that facilitate translocation of a wide variety of substrates across cellular membranes [6,7]. Partanen *et al.* [8] identified polymorphisms in the major histocompatibility complex (MHC) class III gene encoding the G8 protein, where it has been cause a silent mutation.

G8 protein contains about 75 amino acids and contains 10 β -strands and 1 helix in its secondary

structure (<http://prosite.expasy.org/PDOC51484>). These strands are separated by conserved glycine residues and contain some conserved hydrophobic residues. Most G8-containing proteins are predicted to be membrane-integral or secreted. The G8 domain may be involved in extra cellular ligand binding and catalysis. Its predicted structure suggests that it is an integral membrane receptor with extracellular protein-interaction sites and intracellular phosphorylation sites [9] and may interact with extracellular protein-ligands and transduce intracellular signals to the nucleus [10].

The computational methods are used in order to integrate and exploit diverse biological data effectively and elucidate local and genome wide functional connections between protein pairs, and predicted the functional inferences of uncharacterized proteins [11]. In this work, we introduce information-theoretic based approaches to score protein-protein functional interaction, and conserved protein signature matches of uncharacterized protein.

MATERIALS AND METHODS

The uncharacterized G8 protein (*Homo sapiens*) has been selected for this study (Fig. 1) (<http://www.ebi.ac.uk>). This is 75 aa (amino acid) protein and its Accession no are BAB63298.1, GI15277245.

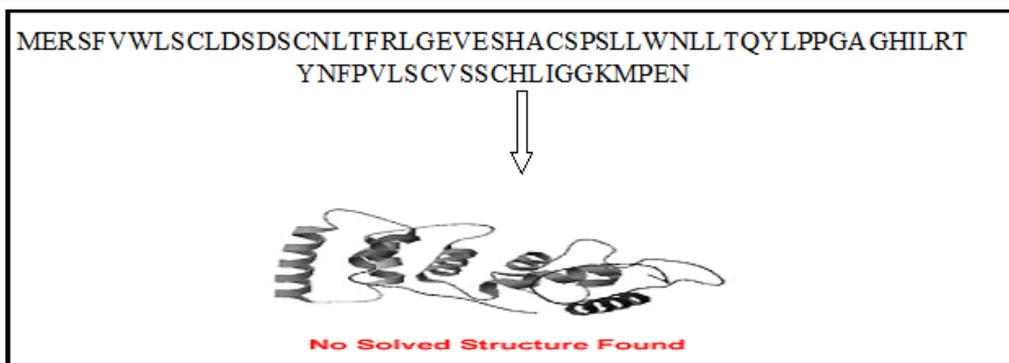


Fig. 1: Three dimensional structure of G8 protein of *Homo sapiens* (Available from <http://www.phosphosite.org>)

For the predicted G8 protein, we first ran BLASTP to search for their homologs [12] from NCBI, using the representative sequences, selected by the CD-HIT program as queries matches. A multiple sequence alignment is built for each cluster using the representative and up to 5 of its nearest sequences by Clustal W (www.genome.jp). The consensus positions where the residues are conserved to $\geq 80\%$ of the sequences. Residues are considered conserved they are identical or belong to the same groups. It ensures that the conservation is maintained by the evolutionary pressure at the protein level.

The evolutionary history of genes helps to predict the functions of uncharacterized genes. The first step is the generation of a phylogenetic tree representing the evolutionary history of the gene of interest and its homologs [13]. The structure of the tree

and the relative phylogenetic positions of genes of different functions are used to trace the history of functional changes, which is then used to predict functions of uncharacterized proteins. The cladograms Guide tree, represent relationships between amino acid sequences to show the relatedness of species.

RESULTS & DISCUSSION

Data base searching, protein sequence pattern analysis, phylogenetic and conserved domain analyses of protein are done to characterize the features of uncharacterized proteins. The results of data base searching indicate that G8 protein is highly conserved among the different organisms such as *Pan troglodytes*, *Nomascus leucogenys*, *Pongo abelii* and *Nomascus leucogenys* indicate good conservation at the amino acid level (Fig 2&3).

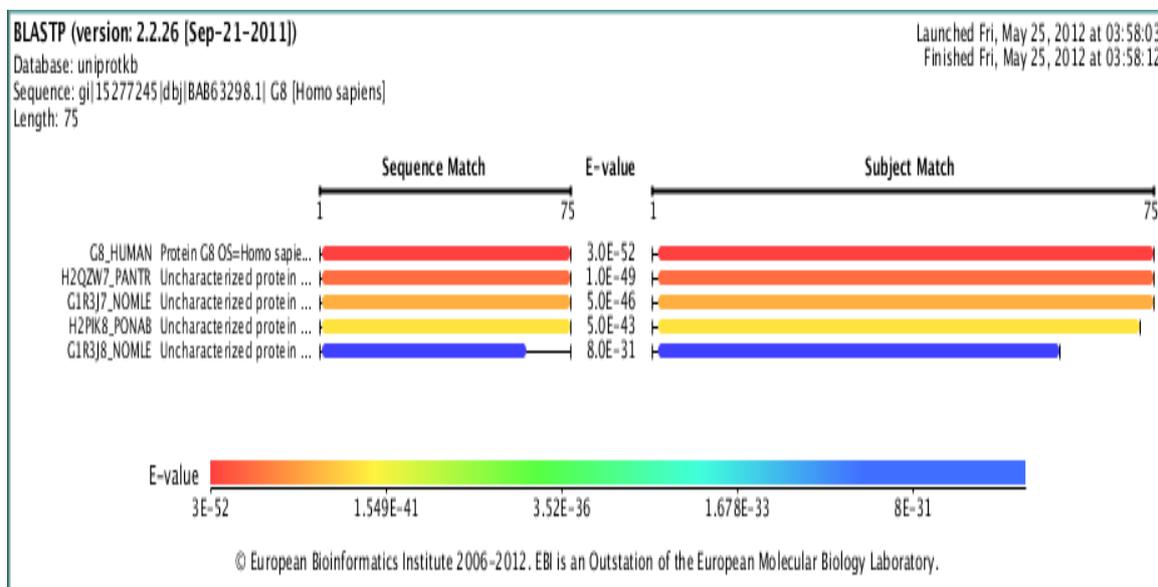


Fig. 2: Visual Blast of G8 proteins of *Homo sapiens* and its homologues (Available from <http://www.ebi.ac.uk>)

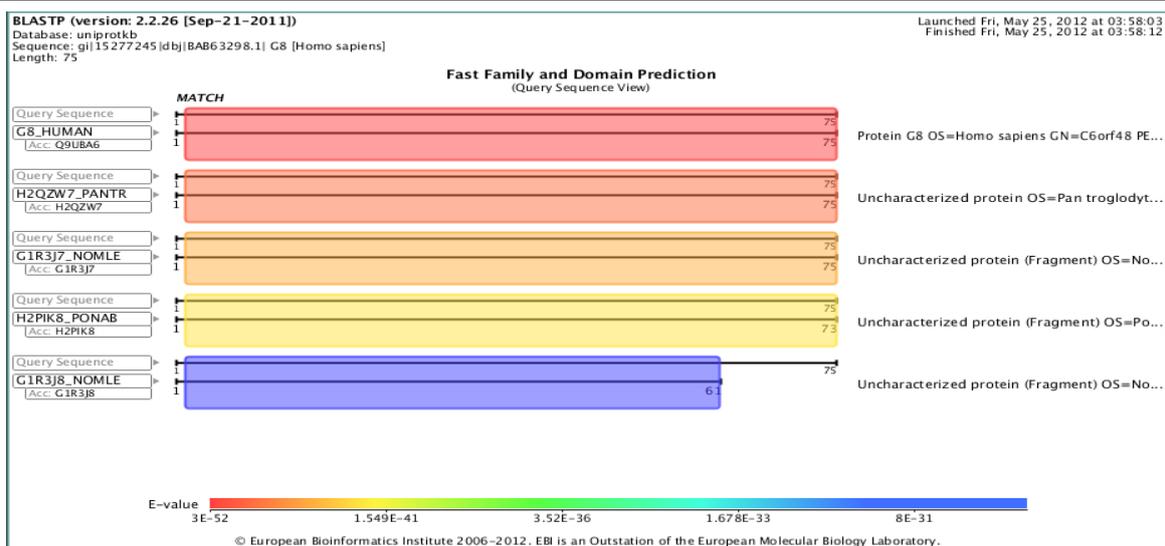


Fig. 3: Fast Family and Domain Prediction of G8 proteins of Homo sapiens and its homologues (Available from <http://www.ebi.ac.uk>)

He *et al.* [14] reported that protein is located at chromosome 6 in *Homo sapiens*, *Pan troglodytes* and *Pongo abelii* but in *Nomascus leucogenys* it is located at Super Contig GL397342.1: 8,645,148-8,650,815. In

all organisms protein is present at forward strand. It is predicted that protein of all organism are closely related to each other, similarly they show the same Cluster272648 (Fig. 4).

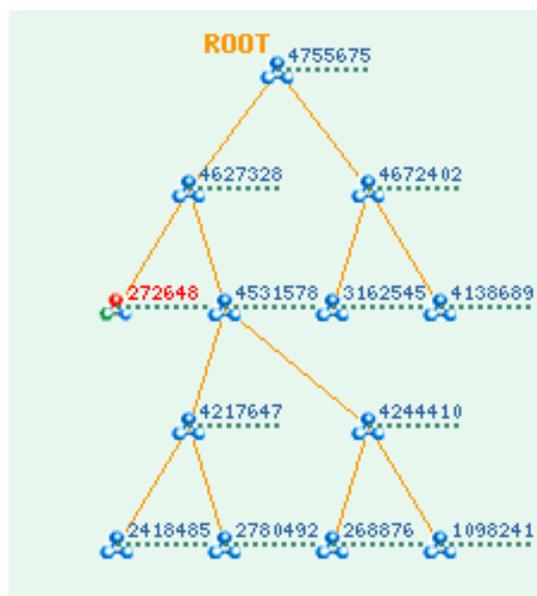


Fig. 4: G8 proteins of Homo sapiens and its homologues showing same cluster 272648

For the G8 protein clusters that have BLASTP homologs, the distribution of the clusters in the main phylogenetic groups & their homologs belong to Eukaryota, The top five dominant classes are only Eukaryota. Result of Proto net (<http://www.protonet.cs.huji.ac.il>) indicate that not only G8 protein of *Homo sapiens* is similar to protein Q9UBA7 but also it is similar to uncharacterized protein of *Pan troglodytes*, *Pongo abelii* and *Nomascus leucogenys*, so it is predicted that all proteins are significantly related to each other.

Multiple sequence alignment of G8 protein and its homologues in different organisms show that these are identical (Fig. 5). The only two exceptions were TR: H2PIK8_PONAB, TR: G1R3J8_NOMLE showed slight variations with G8 protein. Its sequences was 92 % similar to *Pongo abelii* (TR: H2PIK8_PONAB) while in the *Nomascus leucogenys* (TR: G1R3J8_NOMLE) the protein sequence was 83 % similar to G8 protein sequences. The similarity search results against G8 protein sequences were highly significant (based on E-value and score) closely related homologs of the G8 protein [15]. No putative conserved domains have been

detected. The phylogenetic tree showing the evolutionary relationships among various biological

species was based upon similarities and differences in their physical and genetical characteristics (Fig. 6).

```

CLUSTAL 2.1 multiple sequence alignment

sp|Q9UBA6|G8_HUMAN      MERSFVWLSCLDSDSCNLTFRIDGEVESHACSPSLLWNLLTQYLLPGAGHI 50
tr|H2QZW7|H2QZW7_PANTR MERSFVWLSCLDSDSCNLTFRIDGEVESHACSPSLLWNLLTQYLLPGAGHI 50
tr|H2PIK8|H2PIK8_PONAB  VERSFVWLSCLDSDSCNLTFRIDGEVESHACS-SLLWNLLTQYLLPGVGI 49
tr|G1R3J7|G1R3J7_NOMLE  VERSFVWLSCLDSDSCNLTFRILREVESHACSPSLLWNLLTQYLLPGAGYI 50
tr|G1R3J8|G1R3J8_NOMLE  VERSFVWLSCLDSDSCNLTFRILREVESHACSPSLLWYLVFQYLLPGAGYI 50
;*****
;*****

sp|Q9UBA6|G8_HUMAN      LRTYNFPVLSCVSSCHLIGGKMPEN 75
tr|H2QZW7|H2QZW7_PANTR LRTYNFPVLSCVSSCHLIGGKMPEN 75
tr|H2PIK8|H2PIK8_PONAB  LRT-NFPVLPVSSCHLIGGKMPEN 73
tr|G1R3J7|G1R3J7_NOMLE  LRTYSFPVLPVSSCHLIGGRMPEN 75
tr|G1R3J8|G1R3J8_NOMLE  LRTYSFPVI-CV----- 61
*** .***; **
    
```

Fig. 5: Cluster 2.1 multiple sequence alignment of G8 proteins of Homo sapiens and its homologues

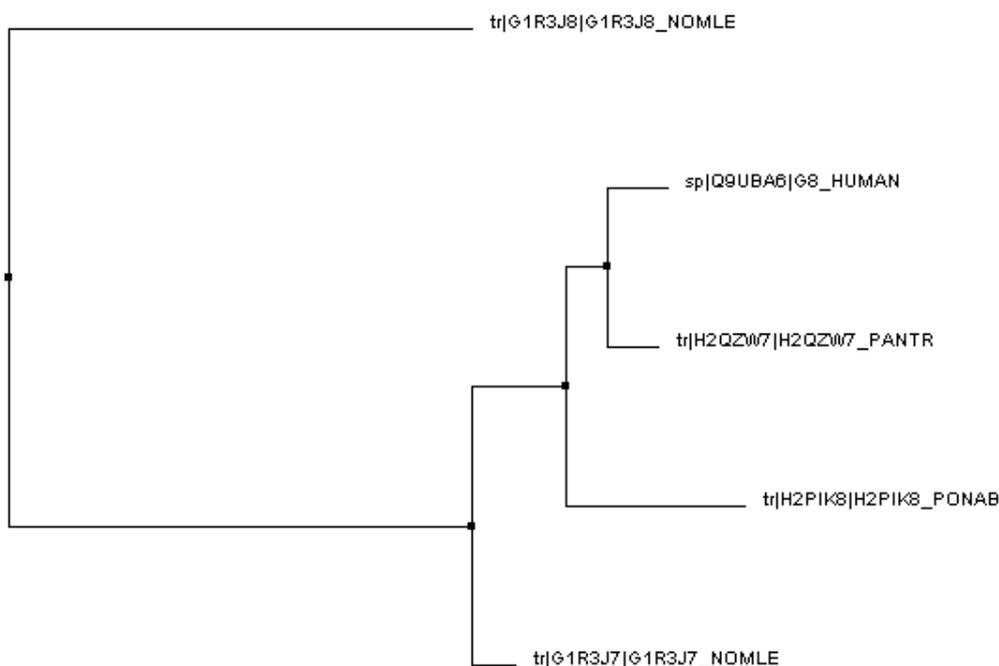


Fig. 6: Neighbor joining tree using of G8 proteins of Homo sapiens and its homologues

CONCLUSION

The Data base searching results indicate that G8 protein of *Homo sapiens* is closely related to *Pan troglodytes*, *Nomascus leucogenys*, *Pongo abelii* and *Nomascus leucogenys*. The Phylogenetic result of G8 protein shows maximum similarity with *Pan troglodytes* and *Nomascus leucogenys*.

ACKNOWLEDGEMENT

I am thankful to C.S.J.M.University, Kanpur for providing lab facility.

REFERENCES

1. Alberts B, Johnson A, Lewis J, Raff M., Roberts K and Walter P. Molecular Biology of the Cell, 4th edition New York: Garland Science; 2002
2. Bergmann C, Senderek J, Windelen E, Küpper F, Middeldorf I, Schneider F et al.; APN (Arbeitsgemeinschaft für Pädiatrische Nephrologie). Clinical consequences of PKHD1 mutations in 164 patients with autosomal-recessive polycystic kidney disease (ARPKD), 2005; Kidney Int., 67(3): 829-848.

3. Gregory A, Graf, Liqing Yu, Wei-Ping Li, Robert G, Pamela L, Tuma CJC, Hobbs HH; ABCG5 and ABCG8 Are Obligate Heterodimers for Protein Trafficking and Biliary Cholesterol Excretion. *J Biol Chem.*, 2003; 278: 48275-48282.
4. Abe S, Usami S, Nakamura Y; Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. *J Hum Genet.*, 2003; 48: 564-570.
5. Guo J, Cheng H, Zhao S, Yu L; GG: a domain involved in phage LTF apparatus and implicated in human MEB and non-syndromic hearing loss diseases. *FEBS Lett.*, 2006; 580: 581-584 .
6. Dean M, Rzhetsky A, Allikmets R; The human ATP-Binding Cassette (ABC) transporter superfamily. *Genome Res.*, 2001; 11: 1156-1166
7. Venkateswaran A, Laffitte BA, Joseph SB, Mak PA, Wilpitz DC, Edwards PA *et al.*; 9 Synthetic LXR ligand inhibits the development of atherosclerosis in mice. *Proc Natl Acad Sci USA*, 97: 12097-12102.
8. Partanen J, Milner C, Campbell RD, Maki M, Lipsanen V, Koskimies S; HLA-linked heat-shock protein 70 (HSP70-2) gene polymorphism and celiac disease. *Tissue Antigens*, 1993; 41: 15-19.
9. Ward CJ, Hogan MC, Rossetti S, Walker D, Sneddon T, Wang X *et al.*; The gene mutated in autosomal recessive kidney disease encodes a large, receptor-like protein. *Nat Genet.*, 2002; 30: 259-269
10. Wilson PD; Polycystic kidney disease. *N Engl J Med.*, 2004; 350:151-164.
11. Mazandu GK, Mulder NJ Scoring Protein Relationships in Functional Interaction Networks Predicted from Sequence Data. 2011 *PLoS ONE* 6(4): e18607. doi:10.1371/journal.pone.0018607
12. Ltschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 1997; 25: 3389-3402.
13. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM; *Molecular systematics*, eds Hillis D.M., Moritz C., Mable B.K. (Sinauer Associates, Sunderland, MA), 1996: 407-514.
14. He QY, Liu XH, Li Q, Studholme DJ, Li XW, Liang SP; G8: a novel domain associated with polycystic kidney disease and non-syndromic hearing loss. *Bioinformatics*, 2006; 22(18): 2189-2191.
15. Sagar M, Saxena P, Pandey N, Singh DB; Modelling transmembrane region in G8 protein of Human and other organisms. *Int J Comput Bioinfo in Silico Model*, 2013; 2(4): 152-158.