International Journal of Basic and Applied Biology
p-ISSN: 2394-5820, e-ISSN: 2349-2539, Volume 3, Issue 2; April-June, 2016, pp. 96-101

http://www.krishisanskriti.org/Publication.html

Deciphering Evolutionarily Conserved Orphan G-protein-coupled Receptors from Homolog Cluster

Aman Chandra Kaushik[1], Deeksha Gautam[2] and Shakti Sahi[3]

[1,2,3]School of Biotechnology Gautam Buddha University Greater Noida, 201308, UP
E-mail: [1]amanbioinfo@gmail.com, [2]deekshabioinfo@gmail.com, [3]shaktis@gbu.ac.in

Abstract—*In various organisms, G-protein coupled receptors are the mostly known cell surface receptors playing vital role among many interacting molecule during signal transduction and metabolic pathways. These are the super largest family which delivers signals from extracellular space to intracellular environment. GPCRs having structural characteristics in terms of possessing seven transmembrane helices with distinct length of intra and extracellular loops. G protein-coupled receptors (GPCRs) are one of the largest protein families of seven transmembrane domain receptors with 800 members in human genome which helping as targets for many drugs. In 1994 Attwood and Findlay categorized the A-F six super families. In the present study, phylogenetic tree for seventy one O-GPCRs was constructed. Four O-GPCRs (GPR139, GPR142, GPR148, and GPR157) which were topologically close to each other with minimum evolutionary distances were identified. Further structures to function approach were used to predict the seven transmembrane domain and active site prediction. These findings will help in the deorphanising of the above mentioned O-GPCRs and understanding of biological structure and function of O-GPCRs belonging to same cluster.*

1. INTRODUCTION

GPCRs being one of the largest protein families in human genome have a characteristic seven alpha helical transmembrane ((7TM) structure. The GPCRs recognize extracellular signals and transduced the inside the cell [1]. Orphan GPCRs (O-GPCRs) are present in various tissues and play an important role in many biological processes [2-4]. They are known to be major drug targets for Type 2 Diabetes and playing crucial role in drug designing [5]. Phylogenetic analysis is an efficient method to find out evolutionary and functional relationships and provides an insight into biological systems [6-7]. The phylogenetic analysis methods have been broadly categorized into distance based and character based methods [8]. Seventy one O-GPCRs were selected for the present study. Multiple sequence alignment for all the sequences was performed to identify conserved regions which provide an insight into the structural and functional aspects [9]. The phylogenetic analysis revealed minimum evolution criterion tree clusters. Out of eight clusters we analyzed the fourth cluster having GPR139, GPR142, GPR148 and GPR157 members which play crucial role in diabetes.

2. METHODOLOGY

Multiple sequence alignments of GPR142 with similar species- Similarities identification was done using BLASTp program of non redundant protein sequence (nr) NCBI Database, then alignment of multiple sequence was done using CLUSTALW of MEGA5 application [10]. Alignment was done in two phase's pairwise alignment and multiple alignments, where pairwise gap opening penalty score was 10 and gap extension penalty score was 0.1. For multiple alignments gap opening penalty score was 10 and gap extension penalty score was 0.2. Gonnet approach was used for protein weight matrix and residue-specific and hydrophilic penalties were applied where gap separation distance was 4 and delay divergent cutoff percentage was 30. Similarly MSA of GPR142 was done with similar PDB structure using above mentioned parameters. Then MSA of GPR142 was done with 71 O-GPCRs sequences using above mentioned parameters.

Phylogenetic analysis of GPR142 with similar species- Phylogenetic tree was constructed using NJ method [11], using MEGA5 application [12]. This method constructs a tree that minimizes the sum of all edge lengths, for the small numbers of taxa in tree, where total number of 100 taxa, total number of sites 272. Poisson model method were applied for all selected taxa, where substitution type was amino acid, rates among sites was uniform, pattern among lineages was homogeneous, complete deletion were applied for missing data.

Then Phylogenetic analysis of GPR142 with similar PDB structure using above mentioned parameters. This method constructs a tree that minimizes the sum of all edge lengths, for the small numbers of taxa in tree, where total number of 37 taxa, total number of sites 257.

Then Phylogenetic analysis of 71 O-GPCRs using above mentioned parameters. This method constructs a tree that

minimizes the sum of all edge lengths, for the small numbers of taxa in tree, where total number of 71 taxa, total number of sites 146, where substitution type was amino acid, rates among sites was uniform, pattern among lineages was homogeneous, complete deletion were applied for missing data.

Predictions of transmembrane domain -Seven transmembrane domains were identified in the GPCR139, GPCR142, GPCR148 and GPCR157, by the TMpred program [13]. TMPred is based on statistical method of TMbase algorithms; prediction is made using many combined weight matrices for scoring of TMPred analysis as well as GPCRHMM [14], GPCRHMM is based on HMM algorithms.

Active Site Prediction: - Active Site prediction of four GPCRs GPR139, GPR142, GPR148 and GPR157 was done using SiteMap Schrodinger suites software [15-16]. SiteMap combined several algorithms for binding site investigation, it provides researchers to find and better characteristics of ligand or protein binding sites.

3. RESULTS AND DISCUSSION

We constructed a comprehensive minimum evolution criterion tree for GPCRs protein sequences in human genome. Our purposes were to use this tree for identification of structure to function relationship of similar GPCRs proteins, which are play crucial role in signal transduction. NJ methods build a tree which is helpful for the minimization of the sum of all edge lengths, for the small numbers of taxa in tree Neighbor Joining methods are likely to be same to the minimum evolution criterion tree, where calculate the matrix Q; find the pair taxa I and j, join and created new central node; calculate the distance of each taxa; calculate the distance of each taxa from outside of new node; and then replace the pair and join nearest node shown in Fig. 1. Further fourth cluster were analyzed out of eight clusters because it shown minimum evolution distance comparatively 71 O-GPCRs. Other clusters also shown important evolution distance but we target only GPR142, which play crucial role as a target in case of Type 2 Diabetes.

Phylogenetic analysis of GPR142 with similar species- Phylogenetic tree was constructed of GPR142 with similar species using NJ method. Calculated distance between human GPR142 proteins with other species of GPR142, where total sum of branch length is 3.359 out of total number of 100 taxa which shown shown minimum evolution distance in Fig. 1.

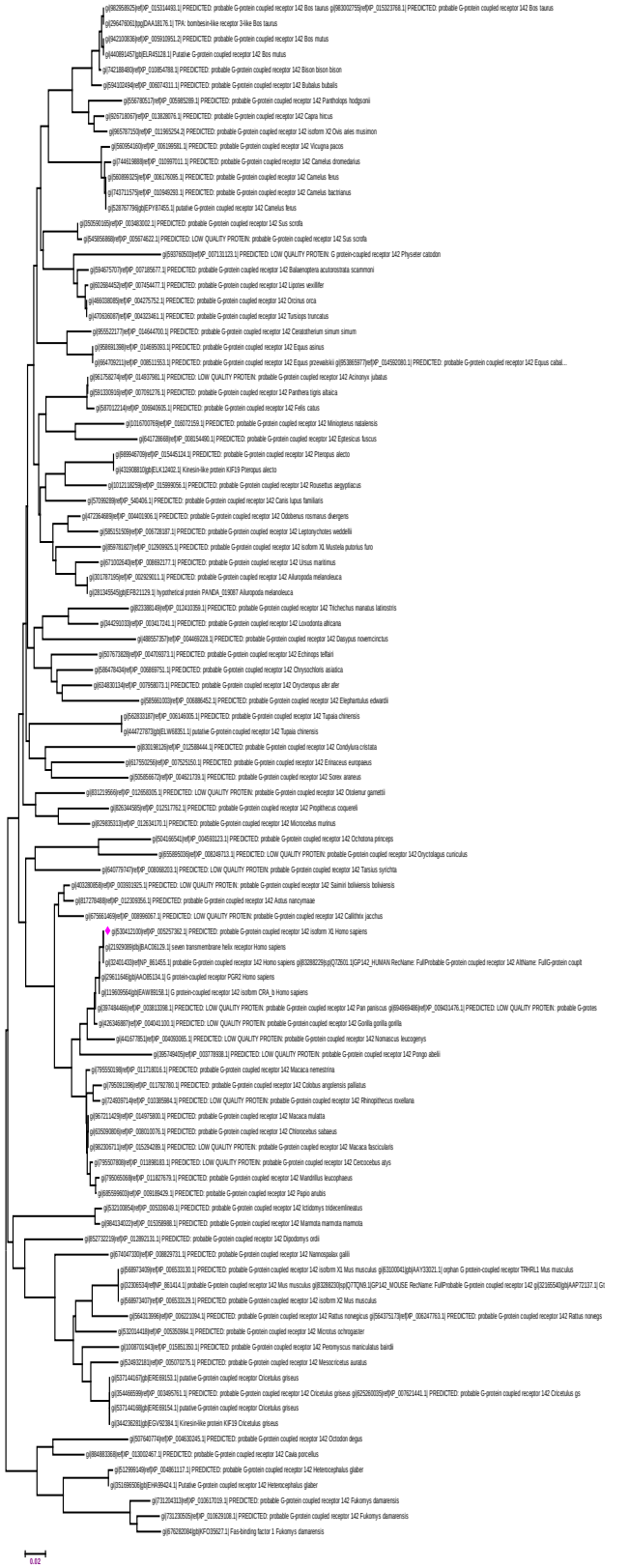

Fig. 1: Phylogenetic tree analysis of GPR142 with similar species using Neighbor joining methods where pink colour indicate the GPR142

Phylogenetic analysis of GPR142 with similar PDB structure- Phylogenetic tree was constructed of GPR142 with similar PDB structure using NJ method. Calculated distance between human GPR142 proteins with other PDB crystal structure available in Protein Data bank (PDB Database) of GPR142, where total sum of branch length is 7.186 out of total number of 37 taxa, which shown minimum evolution distance with other deposited crystal structure available in PDB Database shown in Fig. 2.

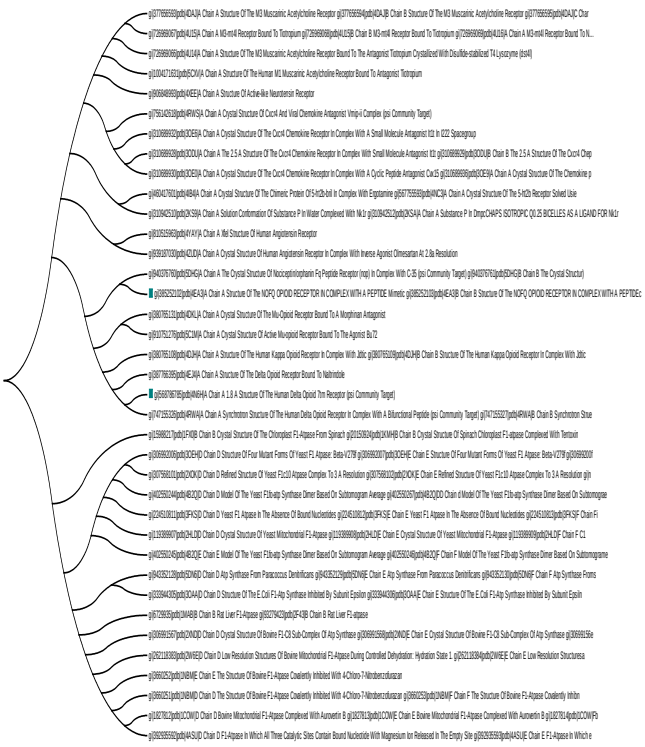

Fig. 2: Phylogenetic tree analysis GPR142 proteins with other PDB crystal structure available in Protein Data bank (PDB Database) of GPR142 using Neighbor joining methods where green colour indicate the similar PDB crystal structure

Phylogenetic analysis of 71 O-GPCRs– Phylogenetic trees were also developed with calculated distance between each GPCRs family, which may be helpful for comparisons of GPCRs with similar domain and function. We covered only seventy one orphan human GPCRs sequences, and collected all the dataset from the IUPHAR database [17]. Jones-Taylor-Thornton model were applied for analysis of statistics where overall distance is 1.727. Statistical model of tree were analysed where total number of params 139, number of rates 1, AICc score is 45176.855, BIC score is 46180.283 and LnL score is -22447.525. Calculated distance between human GPR142 proteins with other orphan GPCRs, where total sum of branch length is 44.302 out of total number of 71 taxa but cluster four shown minimum evolution distance comparatively other cluster shown in Fig. 3.

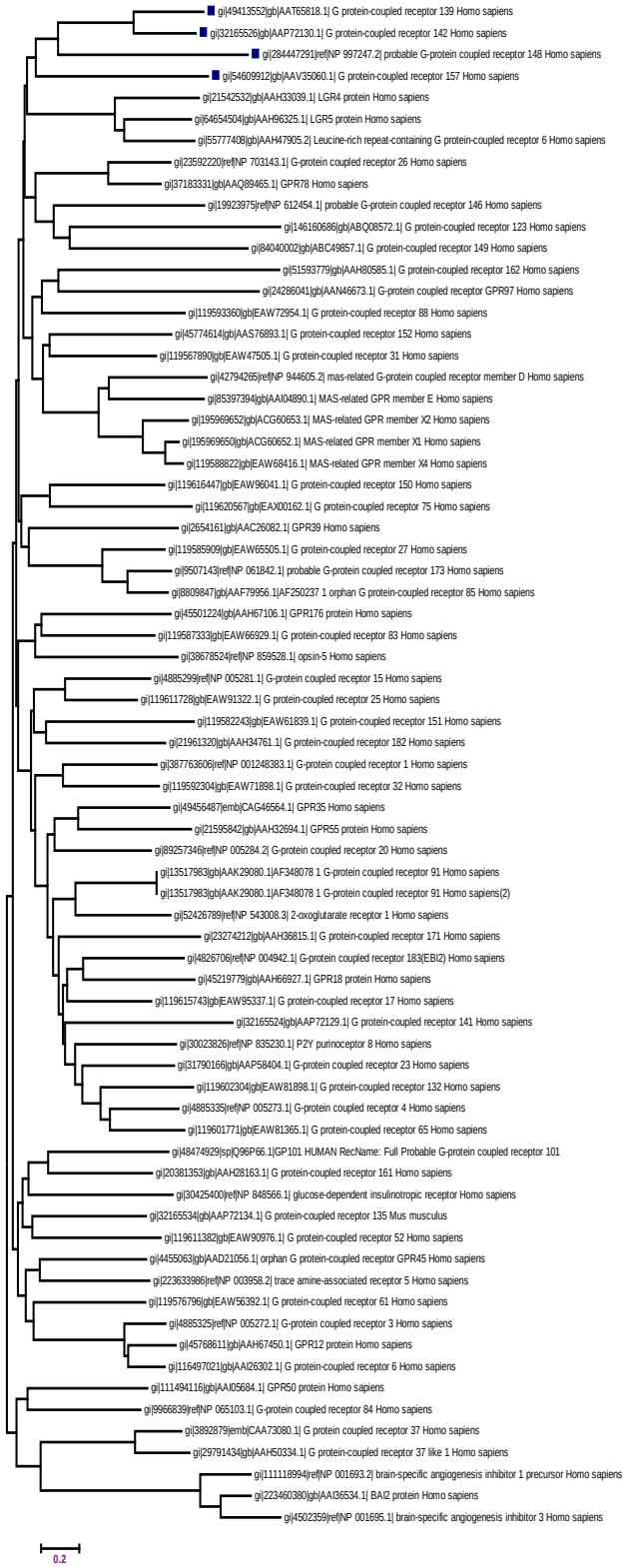

Fig. 3: Phylogenetic tree analysis of GPR142 with other orphan GPCRs using Neighbor joining methods, where blue colour indicate the conserved novel homologues

Conserved Novel homologues- Some cluster shown minimum evolution distance but cluster four shown topologically bare minimum evolution distance, which contain GPR139, GPR142, GPR148, and GPR157 shown in Fig. 4. Fourth cluster was selected which showed minimum score 0.25 between GPR139 and GPR142 and 0.42 score were predicted between GPR149 and GPR157. GPR139 and GPR142 share same ligands for Type 2 diabetes treatment. Sequence of GPR139, GPR142, GPR148 and GPR157 were retrieved from IUPHAR Database, where Sequence format is Pearson Sequence GPE139: gi|49413552|gb|AAT65818.1| which 353 aa long, Sequence GPR142: gi|32165526|gb|AAP72130.1| which 462 aa long, Sequence GPR148: gi|284447291|ref|NP_997247.2| which 347 aa long and Sequence GPR157: gi|54609912|gb|AAV35060.1| which 335 aa long membrane proteins. MSA were done between GPR139, GPR142, GPR148 and GPR157 for validation of similarities shown in Fig. 4.

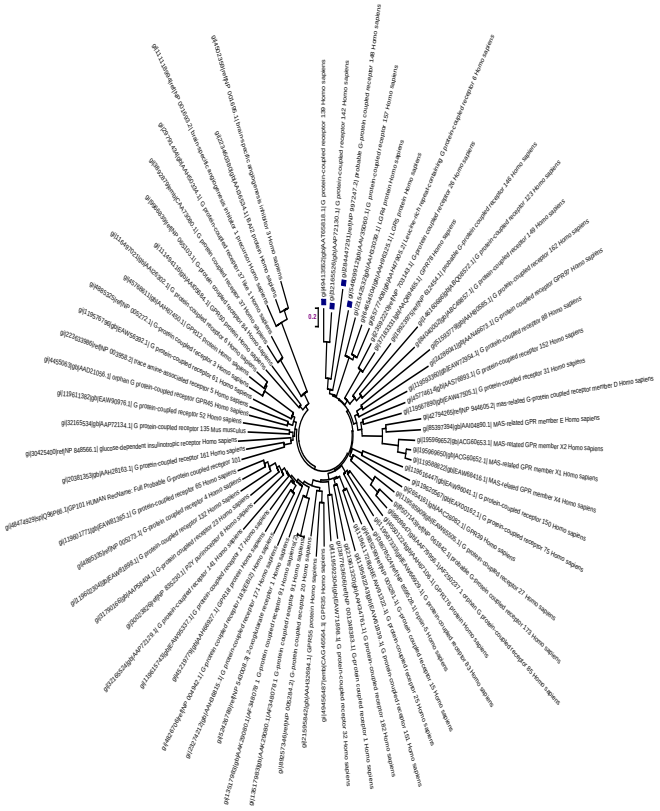

Fig. 4: - Phylogenetic tree cluster using Neighbor joining methods, where GPR139, GPR142, GPR147 and GPR157 shown minimum score show in blue colour

Multiple sequence alignments of Novel homologues- Multiple sequence alignment of GPR139, GPR142, GPR148 and GPR157 was aligned, where parameters were 10 for gap penalty and 0.10 for gap extension penalty, and use BLOSUM30 matrix shown in Fig. 3. GPR139, GPR142, GPR147 and GPR157 are close to each other's and Sequences of (GPR139:GPR142) Aligned. Score: 36.8272, Sequences of (GPR139:GPR148) Aligned. Score: 12.6801, Sequences of (GPR139:GPR157) Aligned. Score: 12.2388, Sequences of (GPR142:GPR148) Aligned. Score: 14.9856, Sequences of (GPR142:GPR157) Aligned. Score: 12.8358, Sequences of (GPR148:GPR157) Aligned. Score: 13.7313 shown in Fig. 5.

1. **GPR139-** GPR139 belongs to Class A family of G Coupled receptor protein, its protein coding gene; chromosome location is 16p12.3 in human. Its seven transmembrane protein and 353 amino acid long receptor protein, and its previous and unofficial name was G(q)-coupled orphan receptor GPRg1, GPRG1 and G-protein-coupled receptor PGR3. Primary transduction mechanisms of GPR139 are act as transducer, and belong to G_q/G_{11} family, which activation of serum response element (SRE) and cAMP response, its inhibitory G-protein and phospholipase C. dimer formation and required proper signaling function in GPCR pathway.
2. **GPR142-** GPR142 belongs to Class A family of G Coupled receptor protein, its protein coding gene; chromosome location is 17q23 in human. Its seven transmembrane proteins and 462 amino acid long receptor proteins, and its previous and unofficial name were KIF19, AXOR103 and G-protein coupled receptor PGR2. Sequence similarities between GPR139 and GPR142 are 50% and it might be share ligands to each others.
3. **GPR148-** GPR148 belongs to Class A family of G Coupled receptor protein, its protein coding gene; chromosome location is 2q14.3 in human. Its seven transmembrane protein and 347 amino acid long receptor protein and previous and unofficial name were BTR, PGR6 and brain and testis restricted GPCR. Tissue distribution of GPR148 is brain and testes in case of human; *GPR148* may be the possible candidate gene, which is involved in developmental delay (DD)/intellectual disability (ID), attention-deficit hyperactivity disorder (ADHD), and neurobehavioral abnormalities due to deletion of chromosome 2q21.1. GPR148 is absent in mouse and rat.
4. **GPR157-** GPR157 belongs to other seven transmembrane family of G Coupled receptor protein, its protein coding gene; chromosome location is 1p36.22 in human, its seven transmembranes protein and 335 amino acids long receptor protein. GPR157 ambiguous sequence similarities to other class A and class B GPCR families, because of distant sequence similarities to other GPCRs, it cannot be classified. GPR139, GPR142, GPR148 and GPR157 identical region (amino acids) shown in dark black colors and similar conserved regions (amino acids) are shown in light black colors.

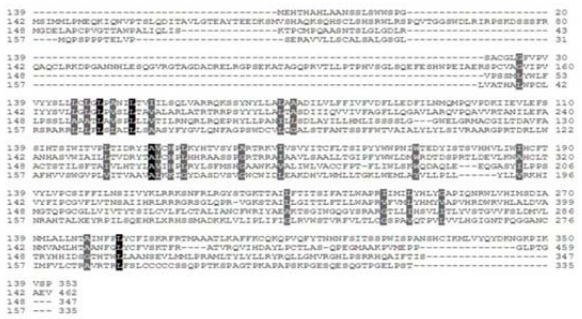

Fig. 5: MSA analysis of GPR139, GPR142, GPR148 and GPR157 GPCRs, where dark black colors shown identical amino acids

Seven Transmembrane Domain analysis- Seven transmembrane domains are identified by the TMpred program and GPCRHMM.

Table1- 7 Transmembrane Domain position prediction of four orphan GPCRs

GPCRs/7TM	GPR139	GPR142	GPR148	GPR157
TM-1	30-51	160-182	51-75	16-39
TM-2	63-83	193-217	90-114	49-69
TM-3	109-129	230-257	127-147	84-106
TM-4	148-167	278-297	168-187	121-142
TM-5	182-208	312-337	215-240	170-192
TM-6	229-253	356-377	261-282	226-248
TM-7	269-289	398-418	299-320	259-282

Active Site Prediction-Active site prediction of four orphans GPCRs with similar PDB ID which is also helpful in active site prediction, that kind of approach called structure to function relationship. While Active site prediction shown in Table 2-

Table 2: Active site prediction of four orphan GPCRs

GPCR	Active Site	Similar PDB
GPR139	24, 43, 47, 51, 54, 55	3NY9
GPR142	88, 89, 92, 93, 167, 172, 216, 220, 223, 250, 254	1GZM
GPR148	89, 92, 93, 184, 234, 238, 266	3AYN
GPR157	59, 151, 152, 162, 233, 237, 251, 252, 255	3PWH

4. CONCLUSION

In this paper, we construct the phylogenetic tree using maximum parsimony and Neighbor joining methods with minimum evolution criterion and calculate distances between seventy one O-GPCRs in human. Some GPCRs show the minimum evolution relationship between some GPCRs (GPR139, GPR142, GPR148 and GPR157), these O-GPCRs grouped in subgroup of taxa. After tree construction and select only GPR139, GPR142, GPR148 and GPR157 for seven transmembrane domains are identified and active site prediction of four O-GPCRs. Subgroup of four O-GPCRs are selected as receptors which belong to different classes, this approach helpful for identification of valuable protein interaction analysis of four O-GPCRs which are closely attached. Systems biology approach will be used to predict the biochemical network.

5. ACKNOWLEDGEMENTS

We thank Dr. Shakti Sahi, Assistant Professor, School of Biotechnology, Gautam Buddha University, India for assistance, guidance and comments that greatly improved the manuscript.

REFERENCES

[1] Pin J.P., Galvez T., Prezeau L.; Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors; Pharmacol. Ther., 98, 2003, 325–354.

[2] McDonald T, Wang R, Bailey W, Xie G, Chen F, Caskey CT, Liu Q., Identification and cloning of an orphan G protein-coupled receptor of the glycoprotein hormone receptor subfamily. Biochem Biophys Res Commun; 247, 1998, 266-70.

[3] Fredriksson R.; The G-protein-coupled receptors in the human genome form five main families: phylogenetic analysis, paralogon groups, and fingerprints; Mol. Pharmacol., 63, 2003, 1256–1272.

[4] Fredriksson R., Schioth H.B.; The repertoire of G-protein-coupled receptors in fully sequenced genomes; Mol. Pharmacol., 67, 2005, 1414–1425.

[5] Kaushik, A. C., & Sahi, S.; Boolean network model for GPR142 against Type 2 diabetes and relative dynamic change ratio analysis using systems and biological circuits approach. Systems and Synthetic Biology, 2015, 1-10.

[6] Gloriam D.E., Schioth H.B., Fredriksson R.; Nine new human Rhodopsin family G protein coupled receptors: identification, sequence characterisation and evolutionary relationship; Biochim. Biophys. Acta, 1722, 2005, 235–246.

[7] Vassilatis DK, Hohmann JG, Zeng H, Li F, Ranchalis JE, Mortrud MT, Brown A, Rodriguez SS, Weller JR, Wright AC, Bergmann JE. The G protein-coupled receptor repertoires of human and mouse. Proceedings of the National Academy of Sciences. 2003 Apr 15;100(8):4903-8.

[8] Bjarnadóttir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, Schiöth HB. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. Genomics. 2006 Sep 30;88(3):263-73.

[9] Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Molecular pharmacology. 2003 Jun 1;63(6):1256-72.

[10] Koichiro T., Daniel P., Nicholas P., Glen S., Masatoshi N., and Sudhir K., MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods; Mol. Biol. Evol. 28(10): 2011, 2731–2739.

[11] Saitou, N., & Nei, M.; The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution, 4(4), 1987, 406-425.

[12] Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in bioinformatics. 2008 Jul 1;9(4):299-306.

[13] Hofmann K. & Stoffel W.; TMbase - A database of membrane spanning proteins segments Biol. Chem. Hoppe-Seyler 1993, 374,166.

[14] Fagerberg L, Jonasson K, von Heijne G, Uhlén M, Berglund L. Prediction of the human membrane proteome. Proteomics. 2010 Mar 1;10(6):1141-9.

[15] Halgren, T., "Identifying and characterizing binding sites and assessing druggability," J. Chem. Inf. Model., 2009, 49, 377–389.

[16] Halgren, T., "New method for fast and accurate binding-site identification and analysis," Chem. Biol. Drug Des., 2007, 69, 146–148.

[17] Harmar, A. J., Hills, R. A., Rosser, E. M., Jones, M., Buneman, O. P., Dunbar, D. R., & Spedding, M.; IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. Nucleic acids research, 37, 2009, 680-685.