Research Article

Homology Modeling of Mouse orphan G-protein coupled receptors Q99MX9 and G2A

Aditya Sood
Assistant Professor, Chandigarh University, Punjab, India

Ridhi Sood
Assistant Professor, Chandigarh University, Punjab, India

Abhijit Kumar
Assistant Professor, Chandigarh University, Punjab, India

Gaganjot Kaur
Assistant Professor, Chandigarh University, Punjab, India

Candy Sidhu
Assistant Professor, Chandigarh University, Punjab, India

Ved Prakash
Department of biotechnology, College of engineering & Technology, IILM-AHL, Greater Noida, India

ABSTRACT

GPCRs are fundamental membrane proteins and they are called as G-Protein Couple Receptors. They are rooted in the cell surface that conveys signals to cell, also known as seven membrane domain receptors. If these receptors do not role correctly, then this may lead to numerous diseases.

Orphan Receptors lack endogenous ligands and their particular purpose is not known. Presently there are few ideal structures in databank for following structural biology studies but no orphan GPCR coordinates exist in protein data bank. The aim for gaining appropriate structures for these proteins is to deorphanize them and to study biological applications. There are numerous GPCRs whose 3D structures are obtainable, using those structures as templates, computer based homology modeling method can be used to presume 3D structures of numerous proteins by means of their amino acid sequence. In this study, we have modeled Q99MX9_MOUSE and G2A_MOUSE, class an orphan GPCRs whose 3D structures have been assumed using Swiss pdb Viewer; the PDB templates used for these orphan GPCRs are 2RH1 and 3UON correspondingly. Based on the outcome of homology modeling of these proteins, their structure has been acquired and their purpose can be presumed on the basis of structural and sequence similarity to known GPCRs. As the structure has been assumed for these GPCRs probable ligands and similarities can be anticipated.

Keywords: GPCRs, Homology Modelling, Structure, Function, Ligand.

Introduction

GPCRs are those proteins which are rooted in the surface of the cell which supports in the transmission of the signals every time there is any kind of reaction to nucleotides, proteins amino acids etc.[1-3] Due to their inappropriate functioning, many diseases arise. This makes GPCRs significant drug targets. The superfamily of GPCRs is the major class of surface receptors of cell, they controls diverse tasks of cell for the physiological responses. This makes GPCRs the significant drug objectives and thus is useful in the pharmaceutical expansions. Numerous of the human derived GPCRs are still orphans who are structure and functions or the ligands have yet not identified.[2-5] This makes GPCR significant to the research related to drug designs. Nuclear magnetic resonance spectroscopy and X-ray crystallography are the main techniques used to recognize the structures of proteins. In NMR spectroscopy the proteins are essential not to be crystallized but dissolved protein with high applications are required. Therefore, GPCR structures cannot be acknowledged with this method. X-ray crystallography is responsible for the atomic information of the proteins which are globular. But GPCRs structures cannot be resolved from this technique because it is very hard to crystallize them.[6,7] Therefore, different structure based methods have to be tracked to resolve the structures and functions of GPCRs. Building models by aligning the target sequences and studying their evolutionary relationships can assistance in formative the structures and functions of GPCRs.[8-10]

Software's and Data sources

- GLIDA: GPCR-Ligand Database
- pBLAST
- Protein data Bank (PDB)
- SWISS-PROT
- Swiss pdb Viewer
- SAVS: Structural Analysis and Verification Server
- Computed Atlas of Surface Topography of proteins (CASTp)

Need of Project

In situation of the human genome, GPCRs are the main gene superfamily. Maximum of the drug based research is going on in this field because not working of these receptors lead to numerous diseases. Considerate their structures, functions and ligand binding sites can aid in drug design which can aid in dealing with the diseases. Deorphanization of orphan GPCRs can widen research and help in discovering new drug targets and also aid in understanding the functioning of the orphan GPCRs on cellular level, as they are involved in numerous signaling pathways.

Methodology

1. Two orphan GPCRs – Q99MX9_mouse and G2A_mouse were selected using GLIDA.

2. The fasta files of these two orphan G-proteins were retrieved.

3. Protein Blast was done and subsequent hits were obtained.

4. The pdb ID's of different templates of interest that are 2RH1 and 3UON were fetched and downloaded using PDB.

5. The fasta sequences of the orphan G-proteins and their respective templates were uploaded in swiss-prot.

6. Homology Modelling was done and the resulting modeled proteins were saved as .pdb files.

7. The energies were checked using swiss pdb viewer. For the structure to be in stable state its energy should be minimum.

8. The ramachandran plots were studied for the modeled proteins and the residues residing in favorable and allowed regions were studied using Structural Analysis and Verification Server.

9. The residues residing in disallowed regions were shifted to allowed regions via loop modeling and

10. Subsequently energies were computed and the modeled structures were checked for the same.

11. The active sites or ligand binding sites were predicted using CASTp server.

12. The fasta sequences of the modeled orphan G proteins were saved and used as an input format in BLAST.

13. Phylogenetic analysis was done.

14. On the basis of Phylogenetic analysis, the possible functions were predicted.

Results and Discussions

All the biological data and the chemical information of the ligands of GPCRs are obtainable in the GPCR-Ligand Database well-known as GLIDA. It includes both bioinformatics and chemo informatics approaches to examine the data of GPCRs [Figure 1-4]. Consequently, it is beneficial in the drug related finding. Mostly the GPCR data is connected to human, mouse and rats. Mouse and rats is the perfect organisms used for the discoveries of numerous drugs. [9-11]

All the data related to ligands such as the active sites, the binding pockets, the charge on the ligands, the volume of the binding pockets (aids in determining the size of ligands) can assistance in calculating the related drugs [Figure 5-7]. Mouse Models are extensively used in drug discovery and clinical trials. If structural and functional information of mouse orphan GPCR can be gained, it can be used in deducing active sites or ligand binding sites for numerous drugs [Figure 8-11]. This is a stepping stone in finding drugs for numerous diseases in humans as mouse and human genetic makeup is very similar. Employing the information about a deorphanized GPCR also will offer useful insights into the purposes related to the protein and various pathways complex in the cells. [10-12]

The Phylogenetic examination aided in understanding and predicting the purposes of modeled proteins. The closely connected proteins display a important use in drug discovery, ligand prediction and additional pharmacological use. Hence we conclude that the modeled G-proteins play a vital role in pharmacological science and predicting specific ligands. [10-12]

REFERENCES

1. Watson S, Arkinstall S. The G protein Linked Receptors Facts Book. New York: Academic Press 1994; 427 p.

2. Flower DR. Modelling G-protein-coupled receptors for drug design. biochimBiophysActa 1999; 1422: 207-234.

3. Takeda S, Kadowaki S,Haga T, Takaesu H,Mitaku S (2002) Identification of G protein coupled receptor genes from the human genome sequence. FEBS Lett 520: 97-101.

4. Collins FS. Finishing the euchromatic sequence of the human genome. Nature 2004; 431: 931-945.

5. LDrews J. Drug Discovery: A historical perspective. Science 2000; 287: 1960-1964

6. George SR, O'Dowd BF, Lee SP. G-protein –coupled receptor oligomerisation and its potential for drug discovery nature Rev. Drud Discov 2002; 1808-820.

7. Skolnick J, Fetrow JS, Kolinski A. Structural genomics and its importance for gene function analysis. Nat Biotechnol 2000; 18:283-287.

8. Kuhlbrandt W, Gouaux. Memberane proteins. CurropinStructBiol 1999; 9: 445-447.

9. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 2000; 289: 739-745.

10. AL Hopkins, CR Groom. Nat. Rev. Drug Discovery 2002; 1 727-730.

11. Wess G, Drug Discovery Today 2002; 4: 533-535.

12. Singh SK, Singh A, Prakash V, C SK. Structure modeling and dynamics driven mutation and phosphorylation analysis of Beta-amyloid peptides. Bioinformation 2014; 10:569-74.

Q99MX9_MOUSE		Query
Gene Name	Gpr84	
Family	Putative / unclassified Class A GPCRs	
GPCRDB	Q99MX9_MOUSE	
UniProt	Q99MX9	
IUPHAR		
Entrez Gene	80910	
KEGG		
Sequence	Q99MX9_MOUSE.fasta	

```
>Q99MX9_MOUSE Orphan G protein-coupled receptor 84.
MWNSSDANFSCYHESVLGYRYFAIIWGVAVAVTGTVGNVLTLLALAIRPKLRTRFNLLIANLTLADLLYCTLLQP
        FSVDTYLHLHWRTGAVFCRIFGLLLFTSNSVSILTLCLIALGRYLLIAHPKLFPQVFSAKGIVLALVGS
        WVVGVTSFAPLWNVFVLVPVVCTCSFDRMRGRPYTTILMGIYFVLGLSSVGVFYCLIHRQVKRAARALD
        QYGLHQASIRSHQVAGTQEAMPGHFQELDSGVASRGPSEGISSEPVSAATTQTLEGDSSEAGGQGIRKA
        AQQIAERSLPEVHRKPRETAGARRATDAPSEFGKVTRMCFAVFLCFALSYIPFLLLNILDARGRAPRVV
        HMVAANLTWLNSCINPVLYAAMNRQFRHAYGSILKRGPQSFRRFH
```

Figure 1: Result for Query Q99MX9_MOUSE.

G2A_MOUSE	
Gene Name	Gpr132,G2a
Family	Putative / unclassified Class A GPCRs
GPCRDB	G2A_MOUSE
UniProt	Q9Z282
IUPHAR	
Entrez Gene	56696
KEGG	
Sequence	G2A_MOUSE.fasta

```
>G2A_MOUSE Lysophosphatidylcholine receptor G2A (G-protein coupled receptor
        132.
MRSEPTNAAGNTTLGVTSVLQSTSVPSSETCHVSYEESRVVLVVVYSAVCLLGLPANCLTAWLTLLQVLQRNVLA
        VYLFCLSLCELLYISTVPLWIIYIQNQHKWNLGPQACKVTAYIFFCNIYISILLLCCISCDRYMAVVYA
        LESRGHRHQRTAVTISACVILLVGLVNYPVFDMKVEKSFCFEPLRMNSKIAGYHYLRFTFGFAIPLGIL
        AFTNHQIFRSIKLSDSLSAAQKNKVKRSAIAVVTIFLVCFAPYHVVLLVKAASFSFYQGDMDAVCAFES
        RLYTVSMVFLCLSTVNSVADPIIYVLGTDHSRQEVSRIHTGWKKWSTKTYVTCSKDSEETHLPTELSNT
        YTFPNPAHPPGSQPAKLGLLCSPERLPEELC
```

Figure 2: Result for G2A_MOUSE

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
3PBL_A	Chain A, Structure Of The Human Dopamine D3 Receptor In Comple	79.0	79.0	97%	8e-16	23%	S
3UON_A	Chain A, Structure Of The Human M2 Muscarinic Acetylcholine Rec	75.9	75.9	92%	8e-15	21%	S
4EJ4_A	Chain A, Structure Of The Delta Opioid Receptor Bound To Naltrind	66.6	66.6	89%	9e-12	22%	S
4DKL_A	Chain A, Crystal Structure Of The Mu-Opioid Receptor Bound To A	65.1	65.1	88%	3e-11	21%	S
4DAJ_A	Chain A, Structure Of The M3 Muscarinic Acetylcholine Receptor >	64.7	64.7	89%	4e-11	20%	S
3D4S_A	Chain A, Cholesterol Bound Form Of Human Beta2 Adrenergic Recer	60.8	60.8	30%	5e-10	32%	S
3SN6_R	Chain R, Crystal Structure Of The Beta2 Adrenergic Receptor-Gs Pi	61.2	97.4	57%	5e-10	32%	S
3KJ6_A	Chain A, Crystal Structure Of A Methylated Beta2 Adrenergic Rece	60.5	97.4	57%	5e-10	32%	S
2RH1_A	Chain A, High Resolution Crystal Structure Of Human B2-Adrenergic	60.8	60.8	30%	6e-10	32%	S
2R4R_A	Chain A, Crystal Structure Of The Human Beta2 Adrenoceptor	60.5	97.4	57%	6e-10	32%	S
3P0G_A	Chain A, Structure Of A Nanobody-Stabilized Active State Of The I	60.8	60.8	30%	6e-10	32%	S

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
2VT4_A	Chain A, Turkey Beta1 Adrenergic Receptor With Stabilising Mutati	32.0	32.0	55%	0.60	22%
3PWH_A	Chain A, Thermostabilised Adenosine A2a Receptor >pdb|3REY|A C	30.8	30.8	73%	1.6	22%
3UON_A	Chain A, Structure Of The Human M2 Muscarinic Acetylcholine Rec	29.6	29.6	8%	3.6	32%
8OHM_A	Chain A, Crystal Structure Of Rna Helicase From Genotype 1b Hepa	29.3	29.3	19%	4.9	32%
1HLL_A	Chain A, Nmr Structure Of T3-I2, A 32 Residue Peptide From The A	26.6	26.6	5%	8.0	52%
1HOF_A	Chain A, Nmr Structure Of T3-I2, A 32 Residue Peptide From The A	26.2	26.2	5%	8.8	52%

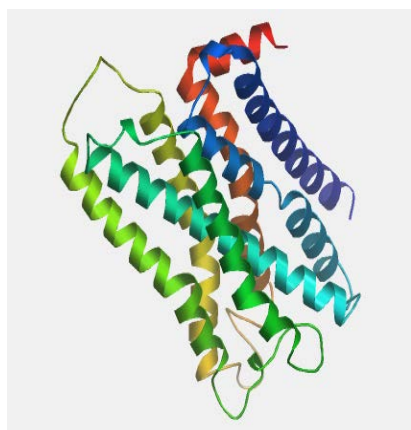**Figure 3:** Blast results

Modeled protein structure of Q99MX9_mouse
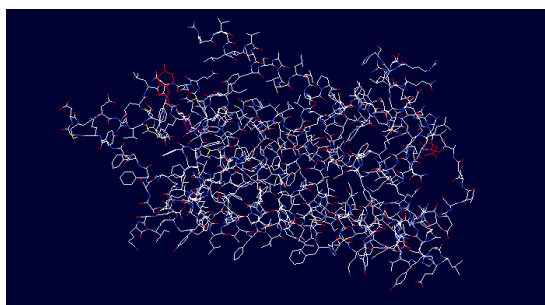

Modeled protein structure of G2A_mouse

Figure 4: Swiss Prot results

(a)

(b)

(c)

Plot statistics

Residues in most favoured regions [A,B,L]	278	87.1%
Residues in additional allowed regions [a,b,l,p]	32	10.0%
Residues in generously allowed regions [~a,~b,~l,~p]	9	2.8%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	319	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	27	
Number of proline residues	16	

Total number of residues	364	

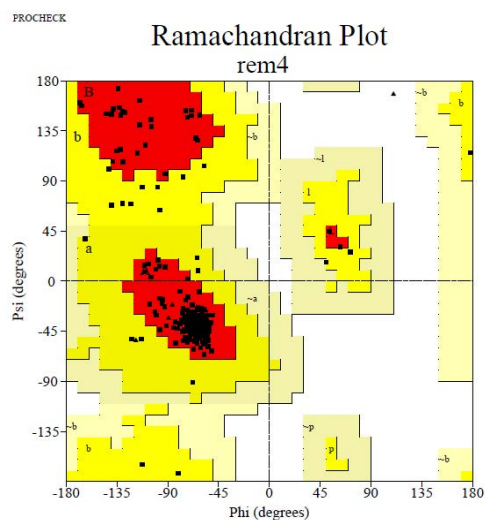| / residue	bonds	angles	torsion	improper	nonBonded	electrostatic	constraint //	TOTAL
KJ/mol	865.676	3054.036	1470.298	675.474	332.79	-10178.81	0.0000 // E=	-3780.534

Figure 5: (a) Protein structure, (b) Ramachandran plot and (c) Energy plot for modeled protein Q99MX9_mouse.

(a)

(b)

(c)

Plot statistics

Residues in most favoured regions [A,B,L]	230	92.0%
Residues in additional allowed regions [a,b,l,p]	20	8.0%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	250	100.0%
Number of end-residues (excl. Gly and Pro)	34	
Number of glycine residues (shown as triangles)	9	
Number of proline residues	8	

Total number of residues	301	

/ residue	bonds	angles	torsion	improper	nonBonded	electrostatic	constraint //	TOTAL
KJ/mol	557.096	2509.668	1007.120	507.632	-1891.67	-7980.61	0.0000 // E=	-5290.770

Figure 6: (a) Protein structure, (b) Ramachandran plot and (c) Energy plot for modeled protein G2A_mouse

AIIWGVAVAVTGTVGNVLTLLALAIRPKLRTRFNLLIANLTLADLLYCTLLQPFSVDTYLHLH
WRTGAVFCRIFGLLLFTSNSVSILTLCLIALGRYLLIAHPKLFPQVFSAKGIVLALVGS
WVVGVTSFAPLWNVFVLVPVVCTCSFDRMRGRPYTTILMGIYFVLGLSSVGVFYCLI
HRQVKRAARALDQYGLHQASIRSHQVAGTQEAMPGHFQELDSGVASRGPSEGISSEP
VSAATTQTLEGDSSEAGGQGIRKAAQQIAERSLPEVHRKPRETAGARRATDAPSEFG
KVTRMCFAVFLCFALSYIPFLLLNILDARGRAPRVVHMVAANLTWLNSCINPVLYAA
MNRQFRHAYGSILK

**Figure 7:** Fasta file of modeled protein Q99MX9_mouse.

AIIWGVAVAVTGTVGNVLTLLALAIRPKLRTRFNLLIANLTLADLLYCTLLQPFSVDTYLHLH
WRTGAVFCRIFGLLLFTSNSVSILTLCLIALGRYLLIAHPKLFPQVFSAKGIVLALVGS
WVVGVTSFAPLWNVFVLVPVVCTCSFDRMRGRPYTTILMGIYFVLGLSSVGVFYCLI
HRQVKRAARALDQYGLHQASIRSHQVAGTQEAMPGHFQELDSGVASRGPSEGISSEP
VSAATTQTLEGDSSEAGGQGIRKAAQQIAERSLPEVHRKPRETAGARRATDAPSEFG
KVTRMCFAVFLCFALSYIPFLLLNILDARGRAPRVVHMVAANLTWLNSCINPVLYAA
MNRQFRHAYGSILK

Figure 8: Fasta file of modeled protein G2A_mouse.
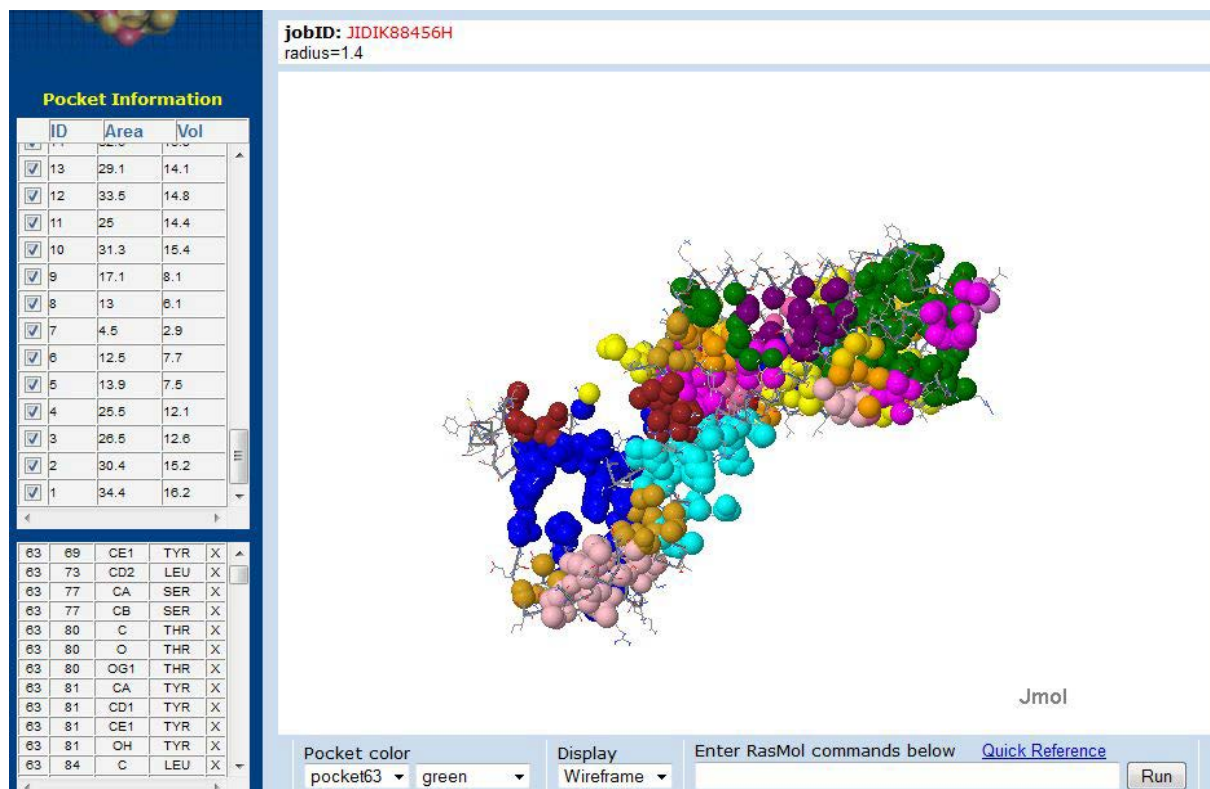
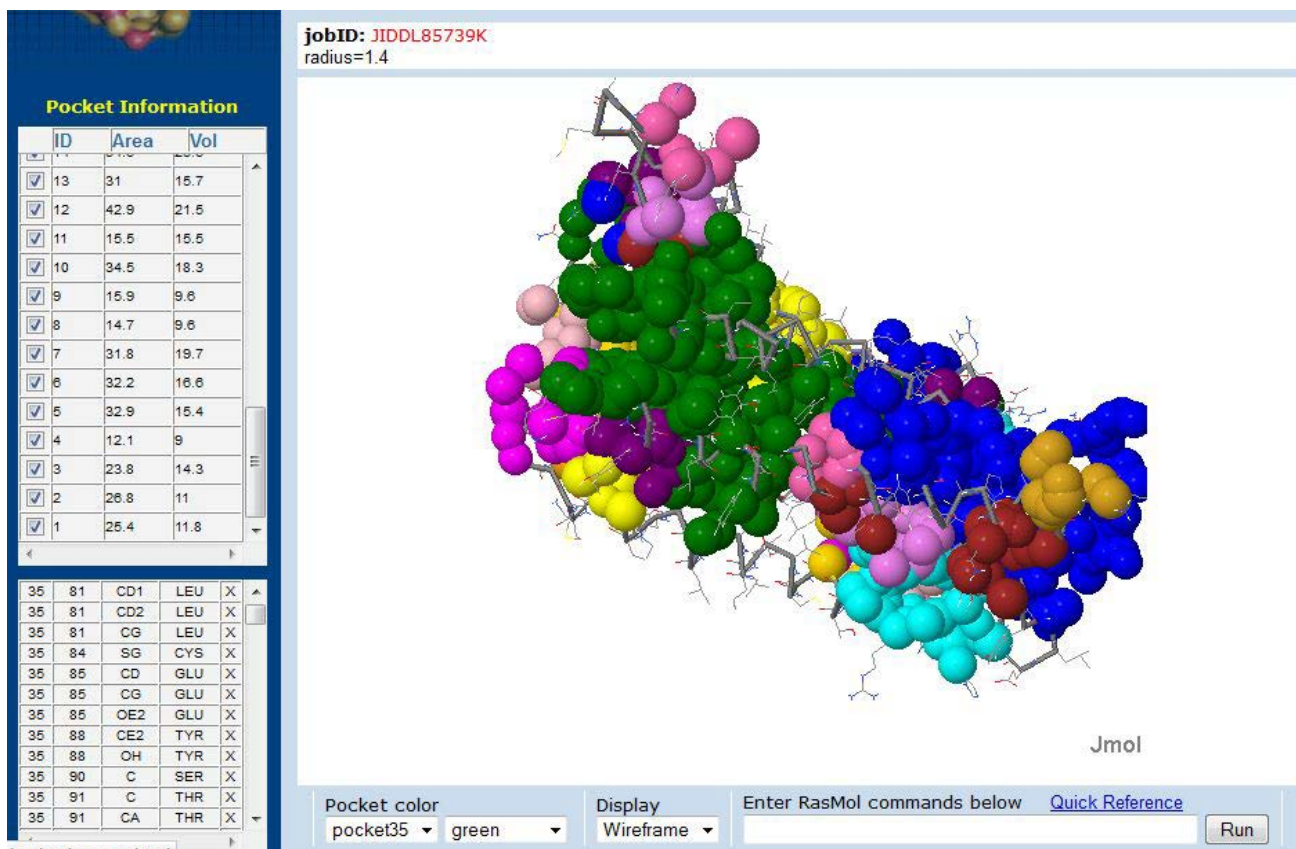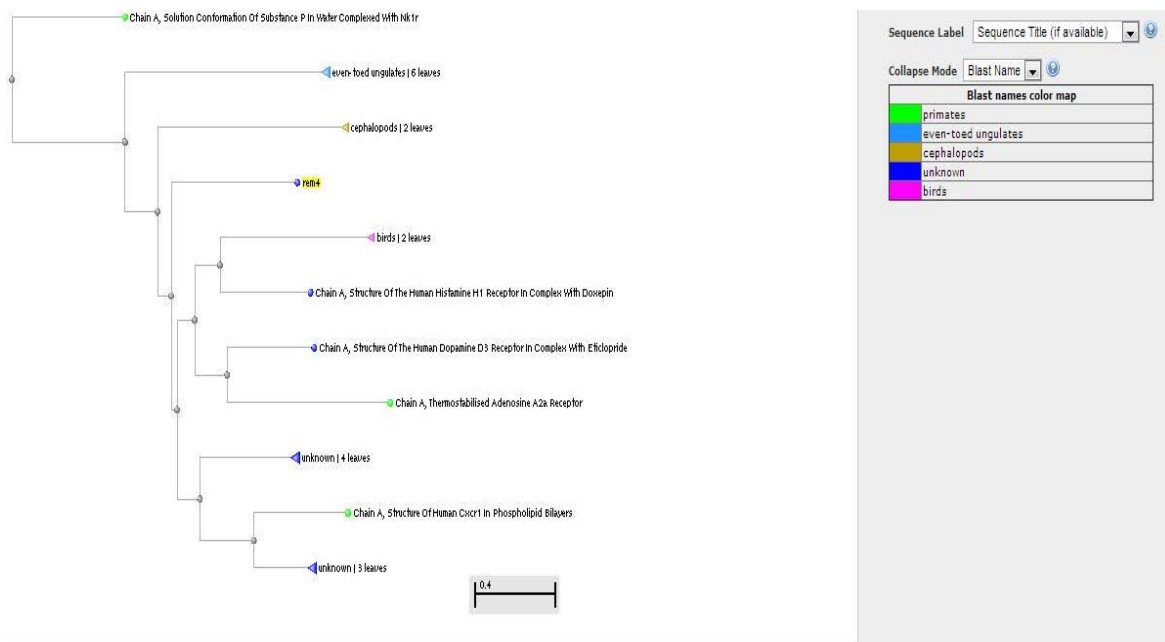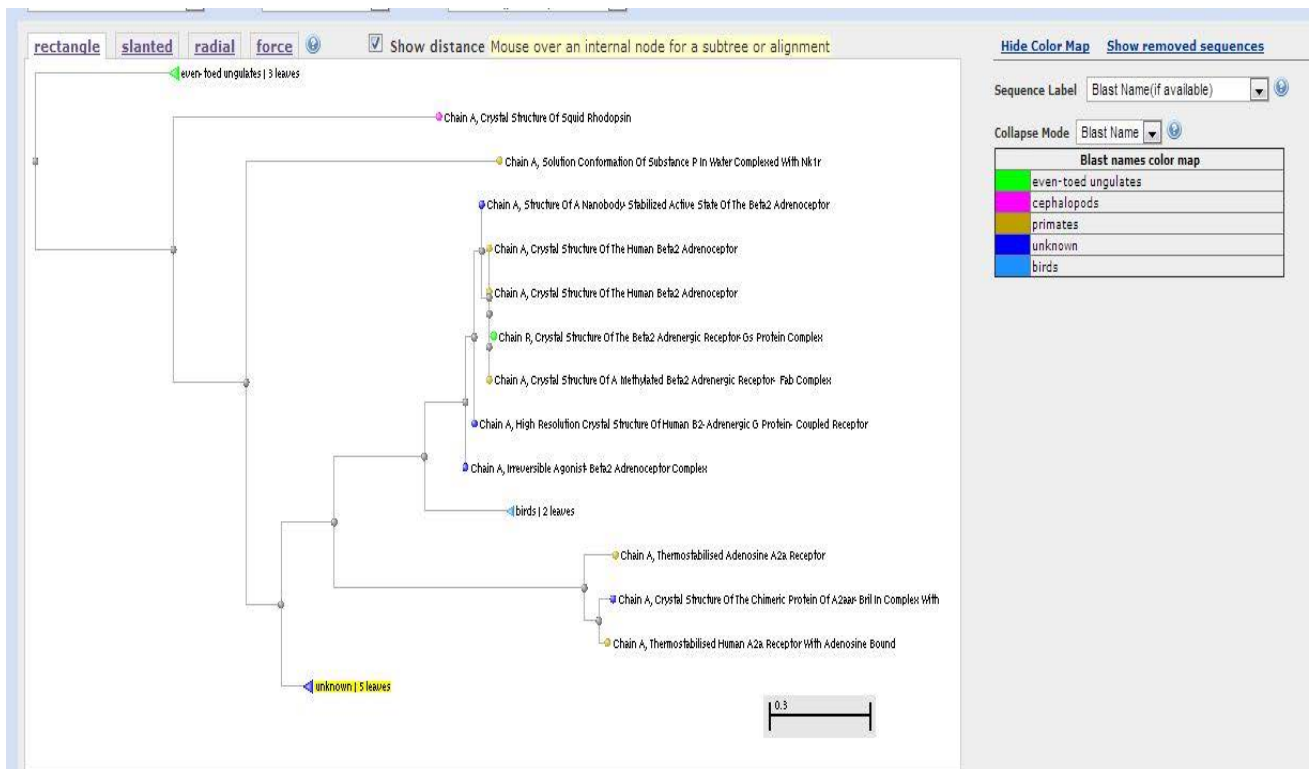**Figure 9:** CASTp results of Q99MX9_mouse.

Figure 10: CASTp results of G2A_mouse.

Figure 11: Phylogenetic Analysis of Modeled Proteins.

ADDRESS FOR CORRESPONDENCE:
Aditya Sood, Department of Preventative Medicines, North-western University Feinberg School of Medicine, Chicago, USA; e-mail: er.adityasood@gmail.com