5th International Conference on Research Developments in Applied Science, Engineering & Management (AEM- 2019)

Mahratta Chamber of Commerce, Industries and Agriculture ,Pune (India) AEM- 2019

23rd February 2019, www.conferenceworld.in ISBN: 978-93-87793-72-9

Homology Modelling and Analysis of GPR3-A Probable GPCR from *Homo sapiens.*

Nazmin Banu C V

Bioinformatics, Indian Institute Of Spices Research, Calicut.

ABSTRACT

Sequencing of the human genome has opened a new era in the life sciences and has greatly accelerated biomedical research. One of the most important areas is the research on orphan G protein coupled receptors (GPCRs). Many intercellular mediators, including peptides, lipids, and other small molecules, have found their GPCRs in the plasma membrane. The aim of this work was to build a three dimensional (3D) model for the protein GPR3 –a probable GPCR from Homo sapiens which was obtained from Swiss-Prot with accession number 'P460F9'. Homology modeling was used for model building. For that we selected 2VT4, 2R4R, 1JFP as templates (structures known) using BLASTP . Three models were generated with these templates using Modeller. For accessing the stereochemical quality of these three models we used PROCHECK.. From our analysis using PROCHECK it is found that the structure (c293) generated with the 2VT4 template suits more for this sequence. The best model was selected based on overall stereochemical quality by Ramachandran plot analysis.

Keywords: Homology modeling, 3D structure prediction, protein 3D structure, GPR3, GPCR

1. INTRODUCTION.

G proteins, short for guanine nucleotide-binding proteins, are a family of proteins involved in second messenger cascades. G proteins are so called because they function as "molecular switches," alternating between an inactive guanosine diphosphate (GDP) and active guanosine triphosphate (GTP) bound state, ultimately going on to regulate downstream cell processes. G protein can refer to two distinct families of proteins. Heterotrimeric G proteins sometimes referred to as the "large" G proteins that are activated by G protein-coupled receptors and made up of alpha, beta, and gamma subunits. There are also "small" G proteins (20-25kDa) that belong to the Ras superfamily of small GTPases. These proteins are homologous to the alpha subunit found in heterotrimers, and are monomeric. However, they also bind GTP and GDP and are involved in signal transduction. G protein-coupled receptors (GPCRs), also known as seven transmembrane domain receptors, 7TM receptors, heptahelical receptors, serpentine receptor, and G protein-linked receptors (GPLR), comprise a large protein family of transmembrane receptors that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. All members of this superfamily share a similar seven transmembrane domain, but are grouped on the basis of shared sequence motifs into different classes such as; Class A

ISBN: 978-93-87793-72-9

Rhodopsin-like, Class B Secretin-like, Class C Metabotropic glutamate/pheromone ,Class D Fungal pheromone, Class E cAMP receptors, Frizzled/Smoothened, Vomeronasal receptors and Putative/unclassified. The pharmaceutical industry has focused on development of modulators to this protein family, but GPCRs also represent an attractive target as biomarkers.

GPCRs are involved in a wide variety of physiological processes including the visual sense, the sense of smell, behavioral and mood regulation, regulation of immune system activity and inflammation, autonomic nervous system transmission and cell density sensing.

The aim of comparative or homology protein structure modeling is to build a three-dimensional (3D) model for a protein of unknown structure (the target) on the basis of sequence similarity to proteins of known structure (the templates). Two conditions must be met to build a useful model. First, the similarity between the target sequence and the template structure must be detectable. Second, a substantially correct alignment between the target sequence and the template structures must be calculated. Comparative modeling is possible because small changes in the protein sequence usually result in small changes in its 3D structure. Although considerable progress has been made in ab initio protein structure prediction, comparative protein structure modeling remains the most accurate prediction method. The overall accuracy of comparative models spans a wide range, from low resolution models with only a correct fold to more accurate models comparable to medium resolution structures determined by crystallography or nuclear magnetic resonance (NMR) spectroscopy. Even low resolution models can be useful in biology because some aspects of function can sometimes be predicted only from the coarse structural features of a model. The 3D structures of proteins in a family are more conserved than their sequences. All current comparative modeling methods consist of four sequential steps; fold assignment and template selection, template–target alignment, model building, and model evaluation. If the model is not satisfactory, template selection, alignment, and model building can be repeated until a satisfactory model is obtained.

G protein-coupled receptor 3, also known as **GPR3**, is a human gene. The protein encoded by this gene is a member of the G protein-coupled receptor family of transmembrane receptors and is involved in signal transduction.

Function:

GPR3 activates adenylate cyclase in the absence of ligand GPR3 is expressed in mammalian oocytes where it maintains meiotic arrest and is thought to be a communication link between oocytes and the surrounding somatic tissue. The orphan receptor GPR3 with constitutive Gs signaling activity is expressed both in peripheral tissues and the CNS. Orphan receptor with constitutive G(s) signaling activity activates cyclic AMP. Has a potential role in modulating a number of brain functions, including behavioral responses to stress, amyloid-beta peptide generation in neurons It also maintains meiotic arrest in oocytes.

2. MATERIALS AND METHODS

2.1 The query protein "**Probable G-protein coupled receptor 3" was taken from UniProt whose accession number** is **P46089**.

Query Protein:

>sp|P46089|GPR3_HUMAN Probable G-protein coupled receptor 3 OS=Homo sapiens GN=GPR3 PE=1 SV=1

MMWGAGSPLAWLSAGSGNVNVSSVGPAEGPTGPAAPLPSPKAWDVVLCISGTLVSCENAL

VVAIIVGTPAFRAPMFLLVGSLAVADLLAGLGLVLHFAAVFCIGSAEMSLVLVGVLAMAF

TASIGSLLAITVDRYLSLYNALTYYSETTVTRTYVMLALVWGGALGLGLLPVLAWNCLDG

LTTCGVVYPLSKNHLVVLAIAFFMVFGIMLQLYAQICRIVCRHAQQIALQRHLLPASHYV

ATRKGIATLAVVLGAFAACWLPFTVYCLLGDAHSPPLYTYLTLLPATYNSMINPIIYAFR

NQDVQKVLWAVCCCCSSSKIPFRSRSPSDV

2.2 Template selection using BLASTP Search against PDB:

This query protein was subjected to BLASTP against PDB for template searching.

Sequences producing significant alignments:		Score (Bits)	E Value
pdb|2VT4|A	Chain A, Turkey Betal Adrenergic Receptor With Sta...	73.6	9e-14
pdb|2R4R|A	Chain A, Crystal Structure Of The Human Beta2 Adre...	62.4	2e-10
pdb|2R4S|A	Chain A, Crystal Structure Of The Human Beta2 Adre...	62.0	3e-10
pdb|2ZIY|A	Chain A, Crystal Structure Of Squid Rhodopsin	51.2	6e-07
pdb|2Z73|A	Chain A, Crystal Structure Of Squid Rhodopsin >pdb...	50.8	6e-07
pdb|3EML|A	Chain A, The 2.6 A Crystal Structure Of A Human A2...	47.8	5e-06
pdb|3C9M|A	Chain A, Structure Of A Mutant Bovine Rhodopsin In...	46.6	1e-05
pdb|2J4Y|A	Chain A, Crystal Structure Of A Rhodopsin Stabiliz...	46.2	1e-05
pdb|1JFP|A	Chain A, Structure Of Bovine Rhodopsin (Dark Adapt...	46.2	2e-05
pdb|2RH1|A	Chain A, High Resolution Crystal Structure Of Huma...	38.5	0.003
pdb|3D4S|A	Chain A, Cholesterol Bound Form Of Human Beta2 Adr...	38.1	0.004
pdb|1HSK|A	Chain A, Crystal Structure Of S. Aureus Murb	29.3	2.2
pdb|1TLP|E	Chain E, Crystallographic Structural Analysis Of P...	28.1	5.1
pdb|1KEI|A	Chain A, Thermolysin (Substrate-Free) >pdb|1KR6|A ...	27.7	5.2
pdb|1GXW|A	Chain A, The 2.2 A Resolution Structure Of Thermol...	27.7	5.3
pdb|1Z9G|E	Chain E, Crystal Structure Analysis Of Thermolysin...	27.7	5.3
pdb|3TMN|E	Chain E, The Binding Of L-Valyl-L-Tryptophan To Cr...	27.7	5.3

Fig 2..2.1 : Descriptive part of BLASTP

Of this we selected 2VT4 (E-value: 9e-14), 2R4R (E-value: 2e-10) and 1JFP (E-value: 2e-05) as templates.

2.3 Alignment using T_COFFEE:

Each of the selected templates was then aligned with the target sequence using T-Coffee.

 ISBN: 978-93-87793-72-9

a. Alignment and score of target and 2VT4 template:

Fig 2.3.1 : Alignment of target and 2VT4 template using TCoffee

b. Alignment and score of target and 2R4R template:

Fig 2.3.2 : Alignment of target and 2R4R template using TCoffee

c. Alignment and score of target and 1JFP template:

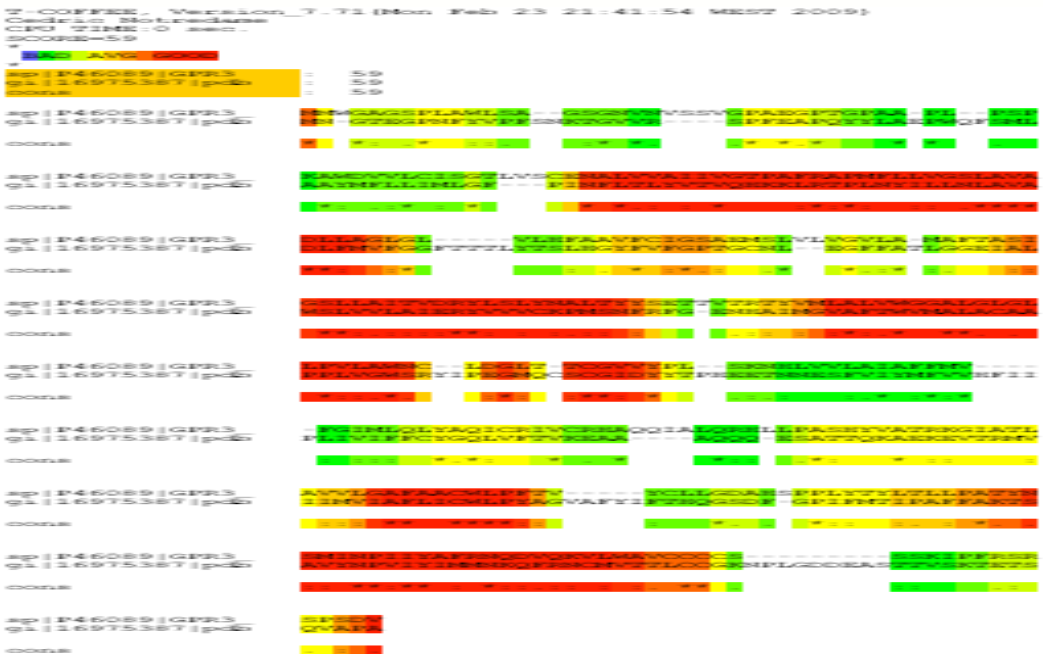


Fig2.3. 3 : Alignment of target and 1JFP template using TCoffee

The following table (Table 2.1) summarizes the result:

Template ID	BLAST e-value of selected templates	T-Coffee Alignment score
2VT4	9e-14	67
2R4R	2e-10	71
1JFP	2e-05	59

Table 2.1

2.4 Modelling using MODELLER:

Each of the Fasta formatted alignment files from T-Coffee were then loaded into Modeller for modelling.

Model 1 (using 2VT4 template):

EXPDTA THEORETICAL MODEL, MODELLER 9v3 2009/04/23 11:07:11

REMARK 6 MODELLER OBJECTIVE FUNCTION: 2348.8440

REMARK 6 MODELLER BEST TEMPLATE % SEQ ID: 23.030

ATOM 1 N MET 1 12.630 4.439 35.828 1.00137.26 1SG 2

 ISBN: 978-93-87793-72-9

ATOM 2 CA MET 1 13.317 3.865 34.649 1.00137.26 1SG 3

……

ATOM 2458 OXT VAL 330 7.178 -5.940 -16.664 1.00 86.33 1SG2459

TER 2459 VAL 330 1SG2460

END

Model 2 (using 2R4R template):

EXPDTA THEORETICAL MODEL, MODELLER 9v3 2009/04/23 11:00:00

REMARK 6 MODELLER OBJECTIVE FUNCTION: 2522.6340

REMARK 6 MODELLER BEST TEMPLATE % SEQ ID: 18.485

ATOM 1 N MET 1 -0.815 -10.367 -51.778 1.00109.08 1SG 2

ATOM 2 CA MET 1 -1.662 -9.559 -52.683 1.00109.08 1SG 3

ATOM 19 CB TRP 3 4.214 -6.251 -53.650 1.00306.81 1SG 20

……

ATOM 2458 OXT VAL 330 55.950 50.111 32.166 1.00 91.56 1SG2459

TER 2459 VAL 330 1SG2460

END

Model 3 (using 1JFP template):

EXPDTA THEORETICAL MODEL, MODELLER 9v3 2009/04/23 11:07:56

REMARK 6 MODELLER OBJECTIVE FUNCTION: 4606.3198

REMARK 6 MODELLER BEST TEMPLATE % SEQ ID: 19.417

ATOM 1 N MET 1 14.965 60.188 58.275 1.00116.07 1SG 2

ATOM 2 CA MET 1 14.134 61.267 58.855 1.00116.07 1SG 3

ATOM 3 CB MET 1 14.445 62.604 58.151 1.00116.07 1SG 4

ATOM 30 O TRP 3 8.126 60.736 54.750 1.00280.32 1SG 31

ATOM 31 N GLY 4 9.351 59.319 53.487 1.00 59.99 1SG 32

……

TER 2459 VAL 330 1SG2460

END

2.5 Model validation using Ramachandran Plot and PROCHECK:

 ISBN: 978-93-87793-72-9

a. Ramachandran Plot of Model 1 (using 2VT4 template) -

c293:

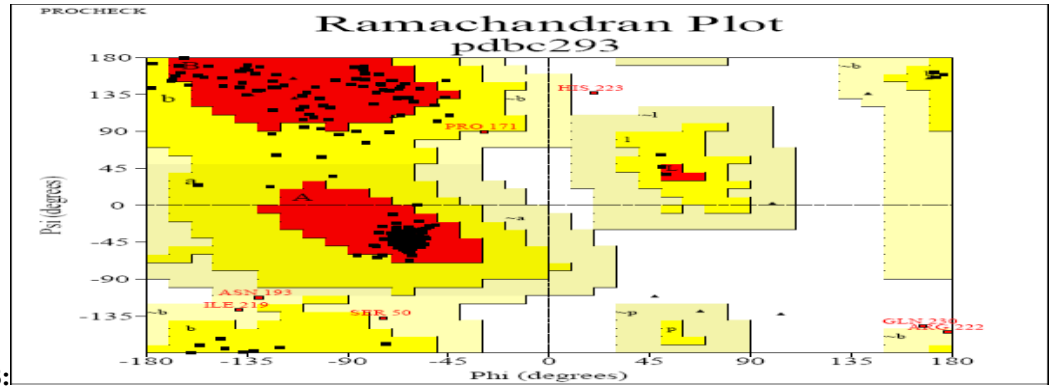


Fig2.5.1 : Ramachandran Plot of c293

Plot Statistics of Model 1 – c293:

Residues in most favoured regions [A,B,L]	244	85.9%
Residues in additional allowed regions [a,b,l,p]	34	12.0%
Residues in generously allowed regions [~a,~b,~l,~p]	4	1.4%
Residues in disallowed regions	2	0.7%
	----	------
Number of non-glycine and non-proline residues	284	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	25	
Number of proline residues	19	

Total number of residues	330	

3D image of structure of PDB code c293:

 ISBN: 978-93-87793-72-9

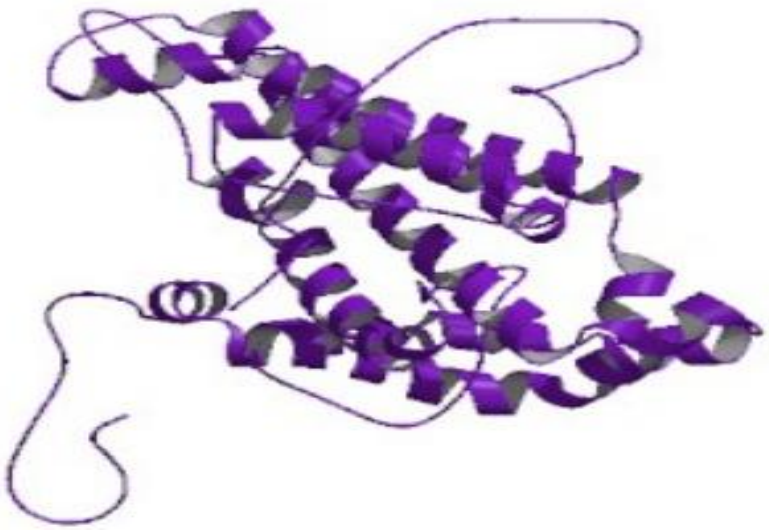


Fig 2.5.2 : 3D image of c293

b. Ramachandran Plot of Model 2 (using 2R4R template) – c301:

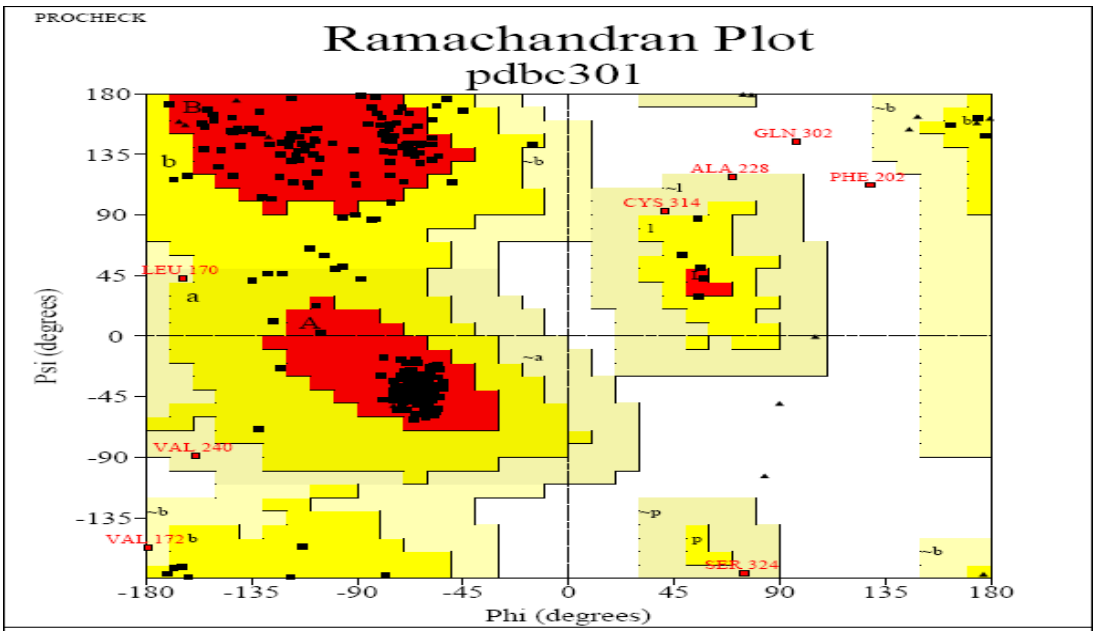


Fig 2.5.3 : Ramachandran Plot of c301

 ISBN: 978-93-87793-72-9

Plot Statistics of Model 2 – c301:

Residues in most favoured regions [A,B,L]	239	84.2%
Residues in additional allowed regions [a,b,l,p]	37	13.0%
Residues in generously allowed regions [~a,~b,~l,~p]	6	2.1%
Residues in disallowed regions	2	0.7%
	----	------
Number of non-glycine and non-proline residues	284	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	25	
Number of proline residues	19	

Total number of residues	330	

3D image of structure of PDB code c301:

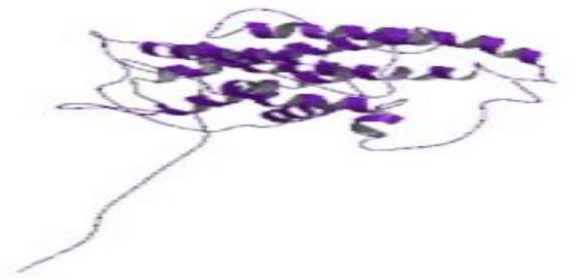
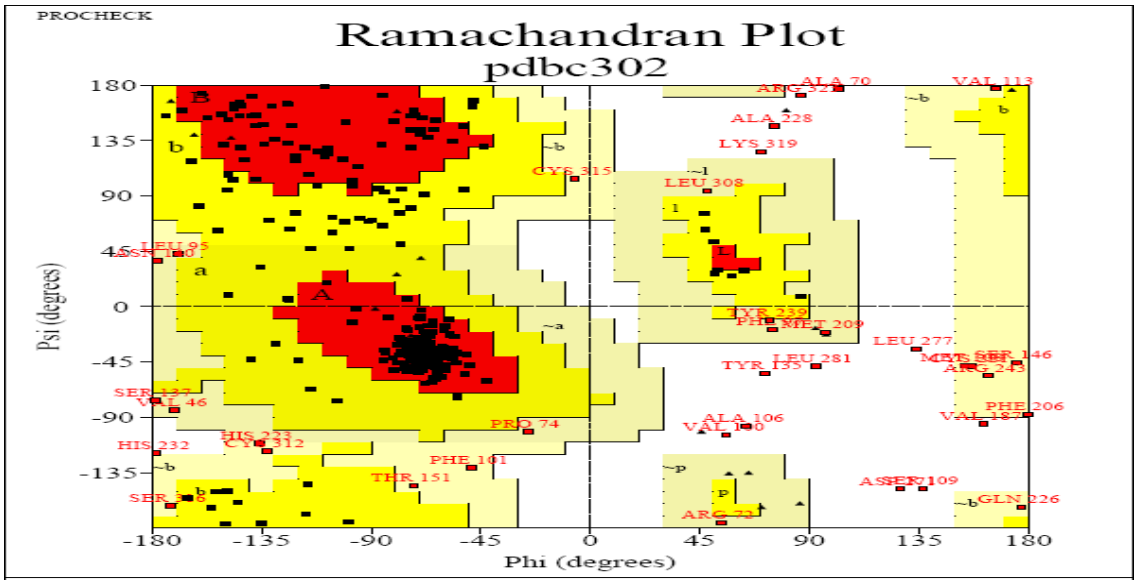


Fig 2.5.4 : 3D image of c301

d. Ramachandran Plot of Model 3 (using 1JFP template) - c302:



ISBN: 978-93-87793-72-9

Fig 3.5.5 : Ramachandran Plot of c302

Plot Statistics of Model 3 –

c302:

Residues in most favoured regions [A,B,L]	192	67.6%
Residues in additional allowed regions [a,b,l,p]	57	20.1%
Residues in generously allowed regions [~a,~b,~l,~p]	22	7.7%
Residues in disallowed regions	13	4.6%
Number of non-glycine and non-proline residues	284	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	25	
Number of proline residues	19	
Total number of residues	330	

3D image of structure of PDB code c302:

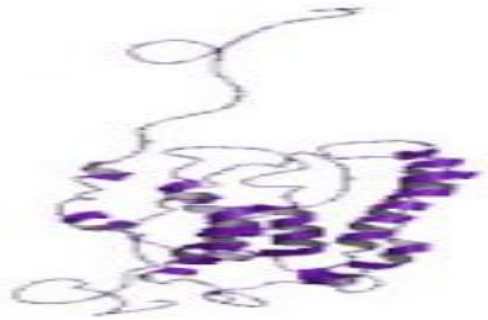
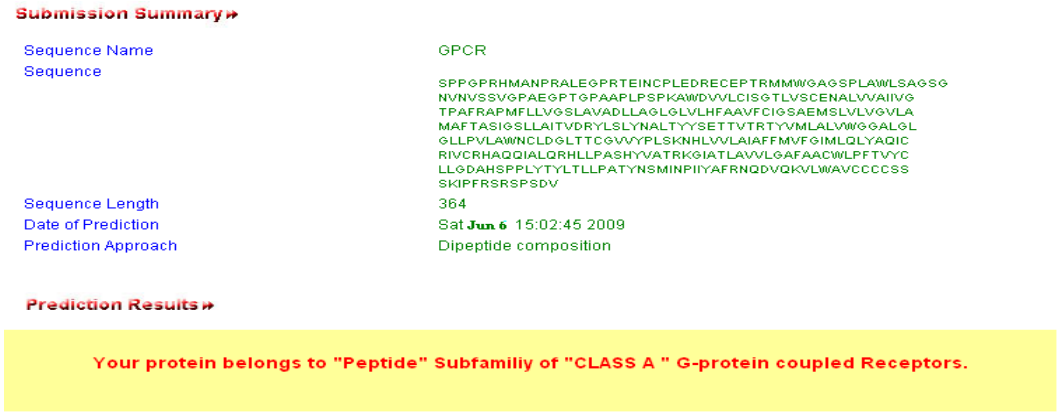


Fig 3.5.6 :3D image of c302

2.6 GPCRpred Result:

To predict which class this GPCR belongs, the FASTA sequence of GPR3 was uploaded to GPCRpred (*http://www.imtech.res.in/raghava/gpcrpred /)* .



Submission Summary»

Sequence Name	GPCR
Sequence	SPPGPRHMANPRALEGPRTEINCPLEDRECEPTRMMWGAGSPLAWLSAGSG NVNVSSVGPAEGPTGPAAPLPSPKAWDVVLCISGTLVSCENALVVAIIVG TPAFRAPMFLLVGSLAVADLLAGLGLVLHFAAVFCIGSAEMSLVLVGVLA MAFTASIGSLLAITVDRYLSLYNALTYYSETTVTRTYVMLALVWGGALGL GLLPVLAWNCLDGLTTCGVVYPLSKNHLVVLAIAFFMVFGIMLQLYAQIC RIVCRHAQQIALQRHLLPASHYVATRKGIATLAVVLGAFAACWLPFTVYC LLGDAHSPPLYTYLTLLPATYNSMINPIIYAFRNQDVQKVLWAVCCCCSS SKIPFRSRSPSDV
Sequence Length	364
Date of Prediction	Sat Jun 6 15:02:45 2009
Prediction Approach	Dipeptide composition

Prediction Results»

Your protein belongs to "Peptide" Subfamiliy of "CLASS A " G-protein coupled Receptors.

Fig 2.6.1 : GPCRpred Result

The result of GPCRpred shows that the query sequence belongs to “Peptide Subfamily” of “CLASS A” GPCRs.

3. CONCLUSION:

Molecular identification of the targets at the DNA level has become possible because of the advancement of techniques in molecular biology, and G protein-coupled receptors (GPCRs) have been found to be the molecular targets of many drugs. The orphan receptor GPR3 is a G protein-coupled receptor endowed with constitutive Gs signaling activity which is expressed both in peripheral tissues and the CNS. GPR3 is involved in several behavioral responses under the control of CNS, in particular the emotional-like responses, Alzheimer’s disease. We predicted a possible structure (c293) for the query sequence (GPR3_Human). Future work such as analysis of the ligand binding pockets of the predicted structure provides valuable insights towards inhibitor design

BIBLIOGRAPHY:

1. Dong-Soon Im D. *Discovery of new G protein-coupled receptors for lipid mediators.* Journal of Lipid Research, 45: 2004 , 410-418
2. Mart´ı-Renom M A, Stuart A C, Fiser A, , S´anchez R, Melo F, and Andrej ˇSali. *Comparative Protein Structure Modelling of Genes and Genomes*. Annual Review of Biophysics and Biomolecular Structure, 29:2000, 291-325
3. Valverde O, Célérier E, Baranyi M, Vanderhaeghen P, Maldonado R, Sperlagh B, Vassart G, and Leden Ct. *GPR3 Receptor, a Novel Actor in the Emotional-Like Responses*. PLoS ONE, 4(3), 2009, e4704.
4. Delaney S, Butler G, Lam C, Thie L,. *Three Improvements to the BLASTP Search of Genome Databases*, , Department of Computer Science, Concordia University

 ISBN: 978-93-87793-72-9

5. Deshpande S K and Sinha R. *Homology modeling of phosphoryl thymidine kinase of enterohemorrhagic Escherichia coli* OH: 157. Bioinformation. 2009; 3(6): 240–243.
6. Chumchua V, Pornputtapong N, Thammarongtham C, and Meksuriyen D. *Homology modeling of Mycoplasma pneumoniae enolase and its molecular interaction with human plasminogen*. Bioinformation. 2008; 3(1): 18–23.