Sequence analysis, structure and binding site prediction of Sigma 1 receptor protein by \textit{in silico} method

Narayanasamy Lokeswaran Latha\textsuperscript{1}, Garimella Gyananath\textsuperscript{1*} & Pudukulathan Kader Zubaidha\textsuperscript{2}

\textsuperscript{1}School of Life Sciences, S. R. T. M. University, Nanded - 431606, India; \textsuperscript{2}Chemical Sciences, S. R. T. M. University, Nanded-431606, India; Garimella Gyananath - Email: gyananath52@gmail.com; Fax: 91-2462-29245; *Corresponding author

Received October 25, 2013; Accepted November 03, 2013; Published December 06, 2013

Abstract:
Sigma 1 Receptor is a subtype of opioid receptor that participates in membrane remodeling and cellular differentiation in the nervous system. Sigma1 Receptor protein with amino acid length ranging from 229 is widely distributed in the liver and moderately in the intestine, kidney, white pulp of the spleen, adrenal gland, brain, placenta and the lung. In this study, the three dimensional structure for sigma 1 receptor protein has been developed by \textit{in- silico} analysis based on evolutionary trace analysis of 37 sigma proteins from different sources. The present work focus on identification of functionally important residues and its interaction with antipsychotic drugs reported in literature.

Key words: Sigma receptor, Evolutionary trace, Binding sites.

Background:
Sigma 1 receptor is a subtype of opioid receptor that participates in membrane remodeling and cellular differentiation in the nervous system. Sigma receptors are subdivided in to 2 subtypes, sigma – 1 and sigma -2. The sigma -1 receptor is a 25K Da protein possessing one putative transmembrane domain and an endoplasmic reticulum retention signal. The cloned sigma-1 receptor reveals no homology to any known mammalian protein \cite{1}. The amino acid sequence of sigma I receptors exhibits more than 90% identity in different mammalian species including guinea pig, human, rat, and mouse \cite{2}. The cloned cDNA, when functionally expressed in mammalian cells, enhances the binding of sigma – 1 receptor ligands. Sigma-1 receptors contain 3 hydrophobic domains namely at the N- and the C termini, and at the center of the protein.

Evolutionary Trace (ET-method) is a method in which protein sequences of a particular protein family is partitioned in to different groups, which originate from the common node in the phylogenic tree and it also involves evolutionary time cut-off. Detailed analysis on the evolutionary conservation information extracted from multiple sequence alignment is an important tool used for prediction of functional properties as well as prediction of ligands binding sites; protein interface surfaces etc, and the detection of conserved residues would be useful in identifying the functionally important residue, even in the absence of structural information. According to the ET method each residue is reported as either conserved or class specific or variable based on the conservation properties \cite{3}. Parthiban \textit{et al} has reported the use of evolutionary trace in identifying conserved and class specific residues close to the putative binding site in nocitcin acetylcholine receptors \cite{4}. LIU Yang Don \textit{et al} \cite{5} has reported the use of evolutionary trace analysis in identifying 11 trace residues in superoxide dismutase of extremophile Thermoplasma acidophilum, of which three residues (Asn39, Gly105 and Glu162) were scattered over the structure and the rest of the residues were identified near the Fe binding site.
In this work, we made an attempt to explore the information about functionally important residues of sigma 1-receptor through evolutionary conservation method and to predict the three-dimensional structure of Sigma 1 protein and its interaction with antipsychotic drugs.

**Methodology:**

**Sequence analysis and ab initio structure prediction of Sigma 1 protein**

The Sigma 1 protein sequence of *Homo sapiens* was retrieved from NCBI (NC_). Domain and pattern were analysed using PFAM [6] and PROSITE [7] database respectively. The secondary structure of the protein was predicted using different JPRED [8], SOPM [9], SOPMA [10] and COR4 [11] servers. The three-dimensional structure of sigma 1 protein was predicted using I-TASSER server (http://zhanglab.ccmbr.med.umich.edu/I-TASSER/) [12]. 3D models are built based on multiple-threading alignments by LOMETS and iterative TASSER assembly simulations; function insights are then derived by matching the predicted models with protein function databases.

**Evolutionary trace analysis**

A total of 74 non-redundant protein sequences of sigma-1 receptor from various organisms were retrieved from Swiss prot database. Of the 74 sequences, only 36 sequences have been selected for multiple sequence alignment and showed in Table 2 (see supplementary material). The selection criteria were based on the (a) sequences having functionally similar domains, (b) sequences with >25% sequence identity, (c) sequences with full length sigma 1 proteins were selected. We performed multiple sequence alignment for 36 Sigma 1 receptor sequences using CLUSTALW. ET analysis for Sigma 1 Receptor sequences were carried out using ET server [13].

**Results & Discussion:**

The 223 amino acids long sigma receptor protein of *Homo sapiens* has an ERG2_Sigma1R. No pattern/signature could be identified from prosite database. No homologous protein with solved structure could be identified from PDB database using BLAST tool. Similarly, threading servers such as PHYRE [14] and 3DPSM [15] also didn’t identify any significant templates that could be used for developing three-dimensional models for sigma 1 receptor protein.

Hence, *ab initio* method was opted for predicting the structure of sigma 1 receptor protein. The protein sequence was submitted to I-TASSER server and five models were predicted. Of the five models, the best model was selected based on TM-score. The C-score and TM-score were -4.12 and 0.28±0.09 respectively. The predicted model was identified to have seven helices and four strands (Figure 1). Based on statistics, if a template/model has a TM-score around or below 0.17, it means the prediction is nothing more than a random selection from PDB library. Hence the predicted model is taken as a significant model.

Based on the evolutionary trace analysis, the phylogenetic tree (Figure 2) was split into 10 evenly distributed partitions, namely P01-P10 in order of evolutionary time cut-off (ETC). Analysis of the mapped traces for partitions P01-P10 revealed clusters of conserved residues occurring on the surface of the protein. Eleven residues were identified to be well conserved in the selected 37 sequences and predicted to be functionally important residues (Figure 3).
Active site prediction of Sigma protein and docking analysis
The predicted sigma protein model was submitted to Q-Site server for prediction of the active site. The first predicted active site of the ten different predictions was selected based on the volume of the site. We then docked the antipsychotic compound (tBOC) into the active site of our target to understand the binding affinity of the compound. Based on the docking results, tBOC binds into the cavity of sigma 1 protein with GOLD score 40.04. The binding of tBOC to sigma 1 protein will be validated using experimental methods. We further docked our compound (tBOC) along with 23 available drug molecules available in PDB database in complex to protein tyrosine phosphatase 1B of Homo sapiens which is an important target for cancer, diabetic and obesity to understand the affinity level of our compound to that of the available molecules. Based on our results Table 1 (see supplementary material), tBOC was ranked sixth among the 23 drug molecules docked to protein tyrosine phosphatase 1B. This result predicts that tBOC has binding affinity with the protein tyrosine phosphatase 1B and can be probably an alternative drug for obesity, diabetics and cancer which has to be further confirmed through wet lab experiments. Further, docking studies were carried out for all these above mentioned 24 drug molecules to sigma 1 protein of human using GOLD software. Based on our results, 2FJ was identified to have a better affinity to sigma 1 protein (GOLD score: 76.0) when compared to tBOC (40.04) (Figure 4).

Conclusion:
Sigma 1 Receptor protein is a potential target for studying basic mechanisms of behavioral studies were shown to be involved in higher ordered brain functions including memory and drug dependence. Since no experimental structures are available for sigma1 receptor protein, our study was first to predict the three dimensional structure of sigma 1 protein. Further we compared the binding of our compound TbOC to sigma 1 protein which proves a potential antipsychotic drug. The predicted functionally important residues and the 3D structure will be helpful for understanding the function of sigma1 receptor protein and for structure based drug designing studies.

References:
[1] Jinbin Xu et al. Nat Commun. 2011 2: 380 [PMID: 21730960]
[2] Hayashi T & Su TP,Curr Neuropharmacol. 2005 3: 267 [PMID: 18369400]
[3] Nagarajan S & Marimuthu P, Bioinformation. 2006 1: 180 [PMID: 17597884]
[4] Parthiban M et al. J Mol Model. 2010 16: 1651 [PMID: 20204665]
[5] Liu Yuan Dong et al. Progress in Modern Biomedicine. 2008 8: 6
[6] Finn RD et al. Nucleic Acids Res. 2010 38: D211 [PMID: 19920124]
[7] Hulo N et al. Nucleic Acids Res. 2006 34: D227 [PMID: 16381852]
[8] Cuff JA et al. Bioinformatics. 1998 14: 892 [PMID: 9927721]
[9] Geourjon C & Deléage G, Protein Eng. 1994 7: 157 [ PMID: 8170919]
BIOINFORMATION

[10] Geourjon C & Deléage G, Comput Appl Biosci. 1995 6: 681 [PMID: 8808585]
[11] Garnier J et al. Methods Enzymol. 1996 266: 540 [PMID: 8743705]
[12] Yang Y, BMC Bioinformatics. 2008 9: 40 [PMID: 18215316]
[13] Innis CA et al. Protein Eng. 2000 13: 839 [PMID: 11239083]
[14] Kelley LA & Sternberg MJE, Nature Protocols. 2009 4: 363 [PMID: 19247286]
[15] Kelley LA et al. J Mol Biol. 2000 299: 499 [PMID: 10860755]

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Edited by P Kangueane

Citation: Latha et al. Bioinformation 9(19): 944-951 (2013)
Supplementary material:

Table 1: List of drug molecules bound with tyrosine phosphatase 1B and their GOLD score.

| S.No | PDB Code | Drug | Structure | Molecule Name | Formula | Gold Score |
|------|----------|------|-----------|---------------|---------|------------|
| 1    | 2VEY     | IZ5  | ![Structure1](image1.png) | N-{[(1S)-2-{[4-{[5S]-1,1-dioxido-3-oxoisothiazolidin-5-yl]phenyl}-1-[4-(3-phenylpropyl)-1H-imidazol-2-yl]ethyl}-3-fluorobenzenesulfonamide | C29 H29 F N4 O5 S2 | 69.07 |
| 2    | 3CWE     | 825  | ![Structure2](image2.png) | [2-bromo-4-{(2R)-3-oxo-2,3-diphenylpropyl}phenyl] (difluoromethyl)phosphonic acid | C22 H18 Br F2 O4 P | 68.99 |
| 3    | 2CNE     | DFJ  | ![Structure3](image3.png) | N-{[(1S)-2-[4-{4-(3-fluorophenyl)methyl}-4,5-dihydro-1H-imidazol-2-yl]ethyl}-3-(trifluoromethyl)benzenesulfonamide | C28 H29 F N4 O5 S2 | 65.02 |
| 4    | 2VEU     | IZ1  | ![Structure4](image4.png) | N-{[1(S)-2-{[5S]-1,1-dioxido-3-oxoisothiazolidin-5-yl]phenyl}-1-[4-phenyl-1H-imidazol-2-yl]ethyl}-3-(trifluoromethyl)benzenesulfonamide | C27 H23 F N4 O5 S2 | 64.34 |
| 5    | 2VEW     | IZ3  | ![Structure5](image5.png) | 3-fluoro-N-{[1(S)-1-[2-fluorophenyl]methyl]imidazol-2-yl}-2-[4-{[5S]-1,1,3-trioxo-1,2-thiazolidin-5-yl]phenyl}ethyl]benzenesulfonamide | C27 H23 F N4 O5 S2 | 60.63 |
| 6    | 2VEX     | IZ4  | ![Structure6](image6.png) | N-{[1(S)-2-{[5S]-1,1-dioxido-3-oxoisothiazolidin-5-yl]phenyl}-1-{[4R]-1-[2-phenylethyl]-4,5-dihydro-1H-imidazol-2-yl]ethyl}-3-fluorobenzenesulfonamide | C28 H29 F N4 O5 S2 | 59.9 |
| 7    | 2NT7     | 902  | ![Structure7](image7.png) | [5-O-][1-(BENZYL SULFONYL)PIPERIDIN-4-YL]AMINO][PHENYL]-4-BROMO-2-(2H-TETRAZOL-5-YL)-3-THIENYL]OXY]ACETIC ACID | C25 H25 Br N6 O5 S2 | 59.9 |
N-[(1S)-1-(4-benzyl-1H-imidazol-2-yl)-2-[(4-[(5S)-1,1-dioxido-3-oxoisothiazolidin-5-yl]phenyl]ethyl]-3-(trifluoromethyl)benzenesulfonamide

N-[(1S)-1-(1H-BENZIMIDAZOL-2-YL)-2-[(4-[(5S)-1,1-DIOXIDO-3-OXOISOTHIAZOLIDIN-5-YL]PHENYL]ETHYL]-4-METHYL-3,4-DIHYDRO-2H,1,4-BENZOXAZINE-7-SULFONAMIDE

N-[(1S)-2-4-[(5R)-1,1-DIOXIDO-3-OXOISOTHIAZOLIDIN-5-YL]PHENYL]-1-[(TRIFLUOROMETHYL)-1H-BENZIMIDAZOL-2-YL]ETHYL]-2,2,2-TRIFLUOROACETAMIDE

4-[3-( dibenzylamino)phenyl]-2,4-dioxobutanoic acid

{5S}-5-[4-[(2S)-2-(1H-BENZIMIDAZOL-2-YL)-2-(1,3-BENZOTHIAZOL-2-YLAMINO)ETHYL]PHENYL]ISOTHIAZOLIDIN-3-ONE,1,1-DIOXIDE

METHYL 2-[[5-[3-CHLORO-4-[(5S)-1,1-DIOXIDO-3-OXOISOTHIAZOLIDIN-5-YL]-N-(PHENYLSULFONYL)-L-PHENYLALANYL]AMINO]PENTYL]OXY]-6-HYDROXYBENZOATE

4-BROMO-3-(CARBOXYMETHOXY)-5-[3-(CYCLOHEXYLAMINO)PHENYL]THIOPHENE-2-CARBOXYLIC ACID
15  2ZN7  410  4-bromo-3-(carboxymethoxy)-5-[3- [cyclohexyl (phenylcarbonyl) amino]phenyl]thiophene-2-carboxylic acid

16  2NTA  521  5-(4-CHLORO-5-PHENYL-3-THIENYL)-1,2,5-THIADIAZOLIDIN-3-ONE 1,1-DIOXIDE

17  2QBP  527  5-(3-[{BENZYSULFONYL}PIPERIDIN-4-YL]AMINO)PHENYL)- 4-BROMO-3- (CARBOXYMETHOXY)THIOPHENE-2- CARBOXYLIC ACID

18  2QBR  910  5-[3-(BENZYLAMINO)PHENYL]-4-BROMO-3-(CARBOXYMETHOXY) THIOPHENE-2-CARBOXYLIC ACID

19  2ZMM  35B  4-bromo-3-(carboxymethoxy)-5-[3- [cyclohexyl (methylcarbamoyl)amino]phenyl]thiophene-2-carboxylic acid

20  2H4G  694  4-BROMO-3-(CARBOXYMETHOXY)-5-(4-HYDROXYPHENYL) THIOPHENE-2-CARBOXYLIC ACID

21  tBoc  Tertiary Butyl Oxycarbony

© 2013 Biomedical Informatics
Table 2: List of selected Sigma 1 receptor sequences with its database accession number and e value which are involved in multiple sequence alignment

| Original S.NO | PDB code NO | Organism       | Accession Number | Domain  | E value |
|--------------|-------------|----------------|------------------|---------|---------|
| 1            | 8188205     | OPRS1_1_RAT    | SP/Q9RC901.1     | EKG2-Σ1r| 1.30E-137|
| 2            | 75040821    | OPRS1_TRIVU    | SP/Q5PX21.1      | EKG2-Σ1r| 1.30E-137|
| 4            | 81862796    | OPRS_1CAVPO    | SP/Q60492.1      | EKG2-Σ1r| 1.20E-125|
| 6            | 11462428    | pann troglodytes| XP-001163744.1   | EKG2-Σ1r| 2.10E-111|
| 10           | 126335022   | Momodophis domistica | XP-001378743.1 | EKG2-Σ1r| 9.00E-114|
| 17           | 140235761   | Neospora laevis | NP-001087013.1   | EKG2-Σ1r| 2.40E-94  |
| 18           | 82083438    | OPR51_TARGR Rec| SP/Q64531.1      | EKG2-Σ1r| 4.00E-101|
| 19           | 56118624    | Xenopus tropicalis| NP-001008207.1   | EKG2-Σ1r| 6.80E-95  |
| 21           | 41055792    | Danio rerio    | NP-957271.1      | EKG2-Σ1r| 2.00E-98  |
| 23           | 82084252    | OPR51_CHEICK   | SPQ5ZI84.1       | EKG2-Σ1r| 2.80E-94  |
| 24           | 72015009    | Stronglycentrotus purpuratus| XP-7838051.| EKG2-Σ1r| 3.10E-91  |
| 25           | 190583472   | Trichoplax adhaerens| gb/EDV23543.1    | EKG2-Σ1r| 8.70E-79  |
| 34           | 156398915   | Nematosella vectensis | XP-001638433.1  | EKG2-Σ1r| 3.10E-64  |
| 29           | 70990744    | Aspergillus fumigatus AI293 | XP-750221.1 | EKG2-Σ1r| 1.80E-131|
| 30           | 119496837   | Neosartorya fischeri NRR1 181 | XP-001265191.2 | EKG2-Σ1r| 3.10E-131|
| 31           | 121702847   | Aspergillus clavatus NRR1 | XP-001269688.1 | EKG2-Σ1r| 1.10E-132|
| 33           | 66862757    | dictyostelium discoideum AX4 | XP-646733.1 | EKG2-Σ1r| 1.00E-33  |
| 34           | 169771107   | Aspergillusoryzae RIB40 | XP-001820023.1  | EKG2-Σ1r| 3.30E-132|
| 35           | 115390859   | Aspergillusiterreus NIH2624 | XP-001212934.1 | EKG2-Σ1r| 3.70E-130|
| 38           | 46125875    | Gibberella zeae FH-1 | XP-387491.1 | EKG2-Σ1r| 5.50E-134|
| 40           | 67516339    | Aspergillus nidulans FGSC a4 | XP-650551.1 | EKG2-Σ1r| 3.70E-133|
| 41           | 156953141   | Sclerotiniacolerotiorum 1980 | XP-001592497.1 | EKG2-Σ1r| 1.50E-116|
| 44           | 14929506    | Lodderomyces elongisporus NRRL YB-4239 | XP-001525629.1 | EKG2-Σ1r| 2.20E-101|
| 45           | 50420709    | Debaryomyces Hansenii CBS767 | XP-458891.1 | EKG2-Σ1r| 3.40E-110|
| 46           | 16442615    | Neurosporacassa OR74A | XP-961312.2 | EKG2-Σ1r| 9.30E-158|
| 47           | 1575320     | C-8 sterol isomerase | GB/ABB94701.0 | EKG2-Σ1r| 3.70E-156|
| 49           | 154318271   | Botryotiniafenchuliana B05.10 | XP-001558454.1 | EKG2-Σ1r| 4.10E-118|
| 50           | 170106151   | Laccaria bicolor S238N-H82 | XP-001884297.1 | EKG2-Σ1r| 2.40E-108|
| 50           | 146413541   | Pichia guilliermondii ATCC 6260 | XP-001482741.1 | EKG2-Σ1r| 1.00E-113|
| 54           | 68473770    | Candida albicans SC5314 | XP-718988.1 | EKG2-Σ1r| 2.10E-106|
| 56           | 126137844   | Pichiaatopsis CBS 6054 | XP-001385445.1 | EKG2-Σ1r| 2.10E-112|
| 58           | 116197971   | Chaetomium globosum CBS 148.51 | XP-001224979.1 | EKG2-Σ1r| 3.30E-127|
| 61           | 50547073    | Yarrowia lipolytica | XP-504668.1 | EKG2-Σ1r| 4.50E-87  |
| 62           | 50554519    | Yarrowialipolytica | XP-501006.1 | EKG2-Σ1r| 2.10E-91  |
| 65           | 19114236    | Schizosaccharomyces pombe 972h-1 | NP-953224.1 | EKG2-Σ1r| 1.20E-153|
| 70           | 45198540    | Ashbya gossypii ATCC 10895 | NP-985309.1 | EKG2-Σ1r| 9.20E-126|
| 74           | 5022117     | Sigma 1 | NP-003857 | EKG2-Σ1r| 6.50E-130|