BMC Bioinformatics

PROCEEDINGS

Open Access

Relationship between amino acid properties and functional parameters in olfactory receptors and discrimination of mutants with enhanced specificity

M Michael Gromiha[1*], K Harini[2], R Sowdhamini[2], Kazuhiko Fukui[3]

Abstract

Background: Olfactory receptors are key components in signal transduction. Mutations in olfactory receptors alter the odor response, which is a fundamental response of organisms to their immediate environment. Understanding the relationship between odorant response and mutations in olfactory receptors is an important problem in bioinformatics and computational biology. In this work, we have systematically analyzed the relationship between various physical, chemical, energetic and conformational properties of amino acid residues, and the change of odor response/compound's potency/half maximal effective concentration (EC50) due to amino acid substitutions.

Results: We observed that both the characteristics of odorant molecule (ligand) and amino acid properties are important for odor response and EC50. Additional information on neighboring and surrounding residues of the mutants enhanced the correlation between amino acid properties and EC50. Further, amino acid properties have been combined systematically using multiple regression techniques and we obtained a correlation of 0.90-0.98 with odor response/EC50 of goldfish, mouse and human olfactory receptors. In addition, we have utilized machine learning methods to discriminate the mutants, which enhance or reduce EC50 values upon mutation and we obtained an accuracy of 93% and 79% for self-consistency and jack-knife tests, respectively.

Conclusions: Our analysis provides deep insights for understanding the odor response of olfactory receptor mutants and the present method could be used for identifying the mutants with enhanced specificity.

Background

Membrane proteins perform several functions, including the transport of ions and molecules across the membrane, binding to small molecules at the extracellular space, recognizing the immune system and energy transducers. Olfactory receptors (OR) are membrane proteins, belonging to the G Protein-Coupled Receptor superfamily, which are characterized by the presence of hydrophobic transmembrane domains. The odorant response of an organism by ORs to its environment forms the basis for our understanding in intra-species interactions, host-pathogen interactions, balance of chemicals, cell-cell interactions and other fundamental processes. It is evident that individual odorant can be recognized by multiple ORs and conversely, one type of OR can recognize multiple odorants with distinct binding affinities and specificities [1,2]. The binding and response of ORs with odorants are critical for the conversion of chemical information into electronic signals in olfactory sensory neurons [3,4]. Recent studies showed that mosquitoes' odorant receptors help the insects to find humans and, inadvertently, to transmit malaria [5,6]. Further, ORs have been analyzed to understand the mechanism of chloride uptake [7], modulation of signaling [8], functional architecture of olfactory system [9], unitary response [10], structural and functional plasticity at binding pocket [11] etc. Similar analysis has also been reported for identifying

* Correspondence: gromiha@iitm.ac.in
[1]Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600 036, Tamilnadu, India
Full list of author information is available at the end of the article

the binding site residues and binding specificity of protein-protein complexes [12-17].

The importance of specific amino acid residues in ORs and other membrane proteins has been demonstrated through site-directed mutagenesis experiments. The experimental data on EC50, maximal velocity of transport, odorant response, percentage uptake of compounds, affinity and specificity have been accumulated in the database for functional residues in membrane proteins [18]. Kuang et al. [19] measured the EC50 values for lysine in the wild type and mutants of 5.24 receptor. Luu et al. [20] elucidated the features of olfactory receptors for determining ligand specificity using different amino acid agonists. The structural basis for mouse OR to EC50 data has been analyzed by systematically substituting amino acid residues in different transmembrane helical segments [2,21]. Schmiedeberg et al. [22] carried out docking studies to understand the influence of different chemical compounds as well as due to mutations. On the other hand, computational methods have been proposed to understand the binding affinity of ligands with ORs using the template structure of rhodopsin [23-25].

In spite of these studies, the role of amino acid properties for the change of EC50 or odorant response has not yet been explored. Further, it is necessary to develop computational models to discriminate the mutants, which increase or decrease EC50. In this work, we have constructed different datasets of goldfish, mouse and human ORs for the mutants that change the odorant response, increase cAMP (adenosine 3'-5'-cyclic mono phosphate) and EC50 values. The differences in experimental data (EC50/odor response etc.) upon mutations have been related with physical, chemical, energetic and conformational properties of amino acid residues and the important amino acid properties have been brought out. The combinations of amino acid properties and the influence of neighboring and surrounding residues have been successfully used to relate the experimental functional data. Further, machine learning methods have been utilized to discriminate mutants with enhanced EC50 values.

Materials and methods
Datasets
We have developed a database, TMFunction, for functionally important amino acid residues in membrane proteins [18]. TMFunction has been searched for all functional data available for ORs. We obtained the experimental data, EC50, odorant response and cAMP increase for goldfish, mouse and human ORs. The final dataset contains 119 data with the following categories: (i) EC50: goldfish OR with Lys, 12; Arg : 12; Gly: 6 and

Glu: 6; mouse OR: 28; Human OR: 7; (ii) odorant response: cAMP increase: 24 and Ca^{2+} increase: 24.

Amino acid properties
We used a set of 49 diverse amino acid properties (physical, chemical, energetic and conformational), which fall into various clusters analyzed by Tommi and Kanehisa [26] in the present study. The amino acid properties were normalized between 0 and 1 using the expression:

$$P_{norm}(i) = [P(i) - P_{min}]/[P_{max} - P_{min}], \qquad (1)$$

where $P(i)$, $P_{norm}(i)$ are, respectively, the original and normalized values of amino acid i for a particular property, and P_{min} and P_{max} are, respectively, the minimum and maximum values. Further, the numerical and normalized values for all the 49 properties used in this study along with their brief descriptions have been explained in our earlier articles [27,28] and are available on the web at http://www.cbrc.jp/~gromiha/fold_rate/property.html. These properties have been successfully used to understand the folding and stability of proteins [29-33].

Computational procedure
The mutation induced changes in property values $\Delta P(i)$ was computed using the equation [29]:

$$\Delta P(i) = P_{mut}(i) - P_{wild}(i) \qquad (2)$$

where $P_{mut}(i)$ - $P_{wild}(i)$ are, respectively, the property value of the ith mutant and wild type residues, and i varies from 1 to N; total number of mutants. The computed difference in property values ΔP (X in Eqn. 3) was related with experimental EC50 or odorant response (Y in Eqn. 3) using single correlation coefficient (r). It is given by:

$$r = [N\sum_{i=1}^{N} X_i Y_i - (\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i)] / \sqrt{[N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2][N\sum_{i=1}^{N} Y_i^2 - (\sum_{i=1}^{N} Y_i)^2]}, \qquad (3)$$

Molecular modelling of mouse OR73
Mouse olfactory receptor 73 (OR73) sequence was obtained from NCBI (http://www.ncbi.nlm.nih.gov/guide/) using text search. The TM regions and topology (N-terminus OUT/IN) of the sequence were predicted using the transmembrane prediction server HMMTOP [34]. The sequences of OR73 and the template (bovine rhodopsin; PDB ID: 1F88A) were aligned using PRALINE-TM [35] server and manually edited using JALVIEW [36] Version 2.4. The TM region predictions from HMMTOP and the two conserved motifs MAYDRYVAIC and NPXXY in OR73 were used to guide and improve the alignment of the query and the template.

The crystal structure of the bovine rhodopsin [23] (PDB ID: 1f88 chain A) was used as a template for the comparative modelling of the query (OR73). The structure of the template was obtained from RCSB (http://www.rcsb.org/pdb). The coordinates corresponding to the residues 236-239 and 328-333 were not available in the crystal structure of 1F88 chain A due to poor electron density and hence these residues were removed from the template sequence before the alignment. The final alignment (Figure 1) was used to construct the model using the software MODELLER [37] (version 9.8). A set of 20 structures were generated from which the five least probability density function models were validated by using PROCHECK [38] (Ramachandran Plot). The best structure was further energy minimized using the SYBYL software package (version 7.2) (Tripos Associates Inc.). Tripos force field, using 100 iterations of Powell's gradient with a distance dependent dielectric constant of 1 and a non bonded interaction cut off value of 8 and was terminated at a convergence of 0.05kcal mol. The final structures were further validated using PROCHECK [38].

Local sequence and structural effects

The effect of local sequence, $P_{seq}(i)$, was included using the equation [29]:

$$p_{seq}(i) = \left[\sum_{j=i-k}^{j=i+k} P_j(i) \right] - P_{mut}(i) \tag{4}$$

where, $P_{mut}(i)$ is the property value of the ith mutant residue and $\Sigma P_j(i)$ is the total property value of the segment of $(2k+1)$ residues ranging from i-k to i+k about the ith residue of wild type.

The structural information, $P_{str}(i)$, was included using the equation:

$$P_{str}(i) = P_{sur}(i) - P_{mut}(i) \tag{5}$$

where $P_{mut}(i)$ is the property value of the ith mutant residue, and:

$$P_{sur}(i) = \sum_j n_{ij} \cdot P_j \tag{6}$$

where, n_{ij} is the total number of type j residues surrounding the ith residue of the protein within the
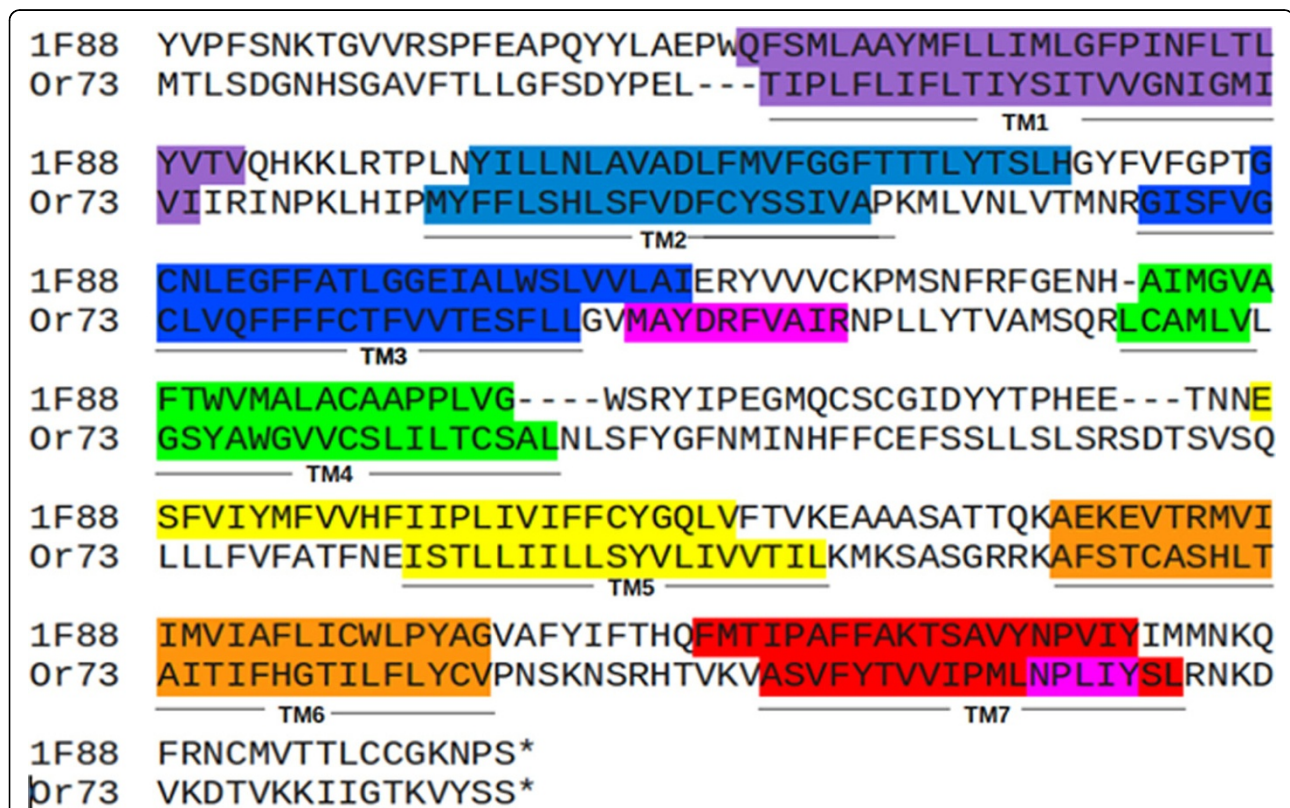

Figure 1 Alignment of query (OR73) with template (1F88). Observed and predicted TM helix positions are marked in VIGBYOR representation. Key functional motifs are shown in pink color against OR73 sequence.

volume of 8Å, and P_j is the property value of residue type j. Further details about the computation of surrounding residues have been described in our earlier articles [39,40].

Multiple regression analysis

We have combined the amino acid properties using multiple regression technique: multiple correlation coefficients and regression equations were determined using standard procedures [41]. When fitting the data by multiple regression technique, reducing the number of variables increases the reliability of results. Hence, we selected three to five properties by searching all possible combinations of the 49 properties and computed the multiple correlation coefficients for all data sets. The highest correlation coefficient was selected and used in the analysis.

Machine learning methods

We have analyzed several machine learning techniques implemented in WEKA program [42] for discriminating mutants with enhanced EC50 values. This program includes several methods based on Bayes function, Neural network, Radial basis function network, Logistic function, Support vector machine, Regression analysis, Nearest neighbor, Meta learning, Decision tree and Rules. The details of all these methods are available in our earlier articles [43-45].

Jack-knife test

We have performed a jack-knife (leave-one-out cross-validation) test for assessing the validity of the present work. In this method, n-1 data are used for training and the left out mutant is used for testing.

Assessment of predictive ability

We have used different measures, such as sensitivity, specificity and accuracy to assess the performance of discriminating mutants with enhanced function. These terms are defined as follows:

$$\text{Sensitivity} = TP/\,(TP + FN) \qquad (7)$$

$$\text{Specificity} = TN/\,(TN + FP) \qquad (8)$$

$$\text{Accuracy} = (TP + TN)\,/\,(TP + TN + FP + FN)\,, \qquad (9)$$

where, TP, FP, TN and FN refer to the number of true positives, false positives, true negatives and false negatives, respectively.

Results and discussion

Relationship between amino acid properties and change in EC50 upon mutation: goldfish OR with Lys potency

We have computed the changes in amino acid properties using Eqn.2 and related them with changes in EC50 upon mutations using Eqn. 3. The accessible surface area of unfolding showed a negative correlation of -0.86 for Lys potency with goldfish OR, and other physical properties, bulkiness, volume etc. showed appreciable negative correlation with ΔEC50 [46]. Figure 2a shows the relationship between ΔASA and ΔEC50. On the
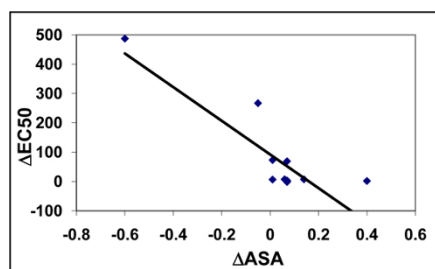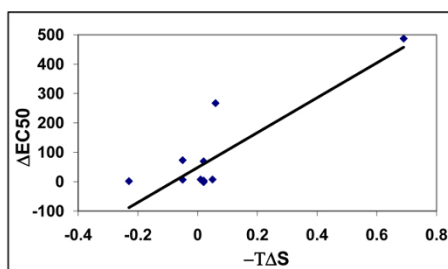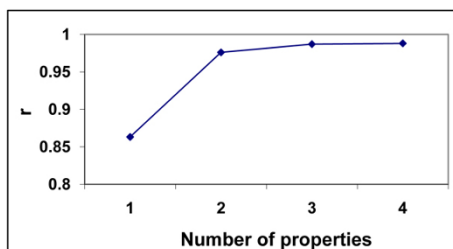

Figure 2 (a) Relationship between ΔASA and ΔEC50. (b) Variation of -TΔS with ΔEC50. (c) Variation of correlation coefficient with number of properties.

other hand, entropy change has high positive correlation with ΔEC50. Figure 2b shows the trend between -TΔS and ΔEC50 and the correlation coefficient is 0.85.

We have analyzed the combined effect of different amino acid properties and related with ΔEC50 values. The variation of correlation coefficient with number of properties is shown in Figure 2c. We noticed that the combination of four properties raised the correlation up to 0.988.

Arg potency

The data on Arg potency with goldfish OR showed a maximum correlation of 0.62 with the property, reduction in accessibility. Further, the combination of five properties also showed the maximum correlation of 0.90. Hence, we have included the information of neighboring residues along the sequence using Eqn. 4. We used different window lengths and the results for window lengths 3, 5 and 7 for the combination of 2, 3, and 4 amino acid properties is shown in Figure 3a.
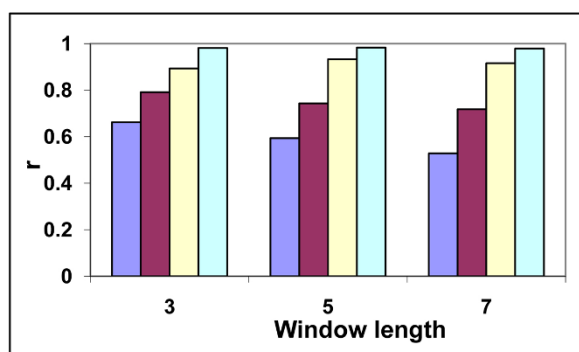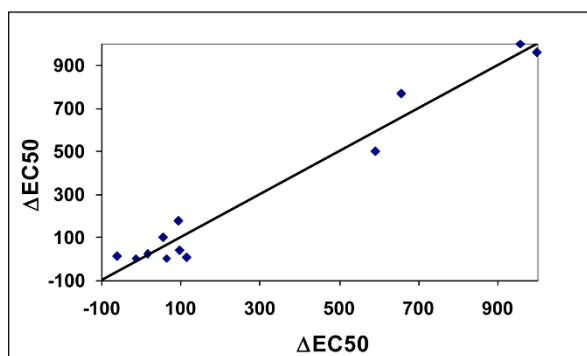
We observed that the combination of four properties raised the correlation up to 0.98. This result shows that the experimental EC50 values are not depending on a specific residue and the information on neighboring residues are very important for the variation of EC50 upon mutation. The experimental and observed ΔEC50 values for all the 12 mutants are shown in Figure 3b and we noticed a good relationship between them. Further, the thermodynamic properties, ΔG, ΔH and -TΔS showed good correlation with ΔEC50 for goldfish OR with Glu and Gly potency.

Molecular modeling and structural analysis of mouse OR

A three-dimensional model of the mouse olfactory receptor, OR73, was obtained as mentioned in Methods. PROCHECK results for the model, excluding the loop regions, show more than 95% of residues in allowed regions (including strictly allowed and partially allowed) of the Ramachandran plot. The full-length structures of the models show more than 90% in the allowed regions
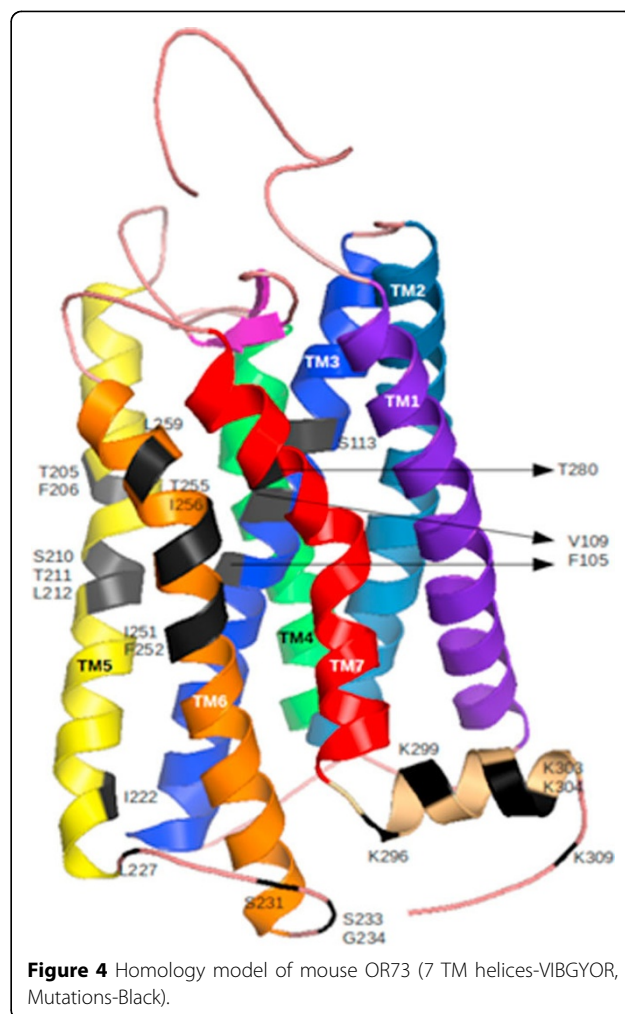

Figure 3a

Figure 3b

Figure 3 (a) Variation of correlation coefficients with different window lengths and combination of amino acid properties: purple: 1 property; brown: 2 properties; yellow: 3 properties and blue: 4 properties. (b) Relationship between experimental and predicted ΔEC50 values.

Figure 4 Homology model of mouse OR73 (7 TM helices-VIBGYOR, Mutations-Black).

(including strictly allowed and partially allowed) of the Ramachandran plot. The residues that were found in the disallowed regions were mainly in the loop regions which are highly variable in length and sequence identity. The RMSD of the full length OR model with respect to template was found to be less than 2Å .The amino acid mutations under study were marked on the model as shown in Figure 4. Many of the mutants lie in TM5 and TM6 and the loop in between has a glycine residue and this region could be implicated in the dimer interface. Most of the other 24 mutants are in intracellular loops.

The solvent accessibility of the amino acid residues, chosen for this study, were observed in the three-dimensional model. These values were calculated using PSA [47] within JOY version 3.2 [48,49] (Additional file 1). It is interesting that only 7 out of 24 mutants are solvent-buried and three of them are in TM3. Such solvent-buried residues could be important for the structural integrity of a protomer.

Relationship between amino acid properties and change in EC50/cAMP increase/Ca^{2+} increase upon mutation: mouse OR

Katada et al. [21] measured the EC50 values for the mutants at various positions in the transmembrane helices of mouse OR. Figure 4 shows a model for mouse OR and the information about mutated residues.

We have computed the difference in amino acid properties and related with difference in EC50 values. We observed a maximum correlation of just 0.38 and the combination of five properties raised the correlation only up to 0.56. We have included the information on neighboring residues, which increased the correlation up to 0.76. We then tried to utilize the structural information of mouse OR using Eqn. 5. The combination of mutants, neighboring residues and surrounding residues enhanced the correlation up to 0.81.

cAMP level increase

Kato et al. [2] measured the increase in cAMP level and Ca^{2+} for 24 mutants, which are located in transmembrane
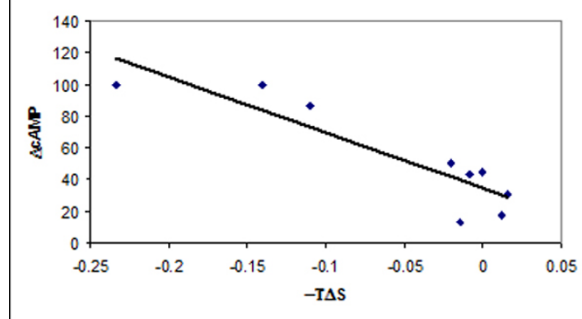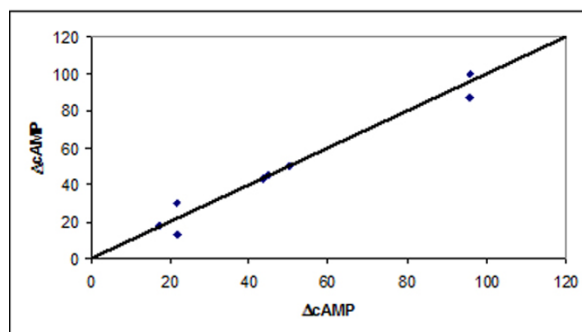

Figure 5 (a) Relationship between -TΔS and cAMP increase in the loop region of mouse OR. (b) Comparison of experimental and predicted cAMP increase levels with the combination of three properties.
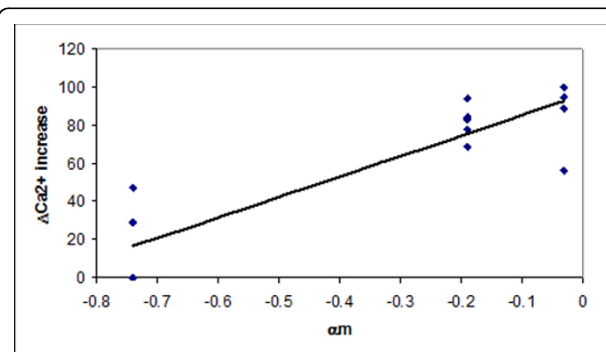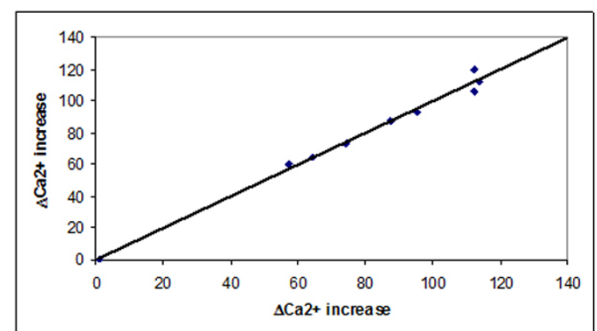

Figure 6 (a) Relationship between α_m and Ca^{2+} increase for the mutants in membrane spanning helical segments. (b) Comparison of experimental and predicted Ca^{2+} increase for the mutants in loop regions.

helical and loop regions. The analysis with overall data did not show high correlation and hence we classified the mutants based on their locations. Interestingly, the classification improved the correlation for the mutants both in transmembrane and loop regions.

We obtained the correlation of 0.86 with the combination of three properties for the mutants in membrane regions. In the loop region, a single property, -TΔS showed a correlation of -0.89 using the information of two neighboring residues on both sides of the mutants (Figure 5a). The combination of three properties raised the correlation up to 0.98. Figure 5b shows the relationship between the experimental and predicted cAMP level increase.

We have analyzed the variation of Ca^{2+} increase upon mutations and we observed that the classification of mutants based on their location is necessary for understanding the properties influencing Ca^{2+} increase. The results obtained for transmembrane helical and loop regions are shown in Figures 6a and 6b, respectively. We obtained a maximum single correlation of 0.90 with the propensity of residues in the middle of α-helix for the mutants in transmembrane helices (Figure 6a). The relationship between experimental and predicted Ca^{2+} increase for the mutants in the loop regions using a combination of three properties is shown in Figure 6b. The correlation coefficient is 0.995.

Relationship between amino acid properties and change in EC50 upon mutation: human OR

Schmiedeberg et al. (2007) measured the EC50 values for seven mutants in human OR. We found that bulkiness has a correlation of 0.71 and the information on three neighboring residues increased the correlation up to 0.92 (Figures 7a and 7b). This analysis reveals the importance of neighboring residues for determining the change in EC50. Further, the combination of just two properties increased the correlation up to 0.999 and Figure 7c showed the relationship between experimental and predicted ΔEC50.

Discrimination of mutants that enhanced/decreased EC50: mouse OR

We have collected a set of 28 data in mouse OR in which 15 of them increased EC50 upon mutation and 13 mutants decreased the EC50 values. We made an attempt to discriminate these mutants using the information on wild type residue, mutant residue and the properties of neighboring residues. We have utilized
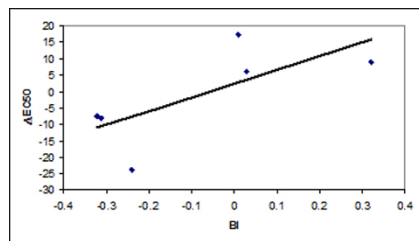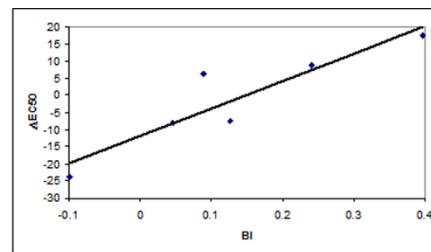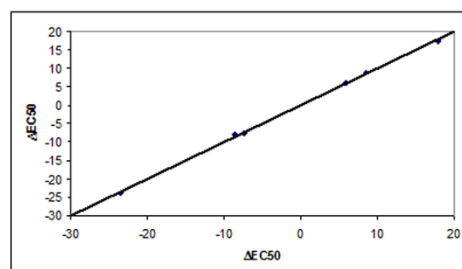

Figure 7a

Figure 7b

Figure 7c

Figure 7 Relationship between bulkiness and ΔEC50 in human OR; (a) Mutant information and (b) Three neighboring residues (c) Comparison of experimental and predicted ΔEC50.

several machine learning techniques for discrimination. Our method showed an accuracy of 92.9% for self-consistency test and the sensitivity and specificity are 93.3% and 92.3%, respectively. The assessment using jack-knife test showed an accuracy of 78.6% and the sensitivity and specificity are 80.0% and 76.9%, respectively. This method can be used to identify the mutants with increased/decreased EC50.

Conclusions

We have constructed different datasets for mouse, goldfish and human ORs and various experimental data such as EC50, odorant response, Ca^{2+} increase etc. The experimental data have been systematically analyzed with physical, chemical, energetic and conformational properties of amino acid residues and important properties have been listed out. We found that the information on neighboring and surrounding residues, namely the inclusion of motifs, is important to understand the function. Further, we have combined the amino acid properties using multiple regression analysis, which relates experimental EC50/cAMP level increase and Ca^{2+} increase very well. We have utilized machine learning techniques for discriminating the mutants that enhance/reduce IC50 upon mutation and we obtained an accuracy of 93% and 79%, respectively, for self-consistency and jack-knife tests. The results obtained in the present work would help to understand the importance of amino acid properties to the functions of ORs and to identify the mutants with enhanced EC50 values.

Additional material

Additional file 1: Solvent accessibility of amino acids under study.

Author details

[1]Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600 036, Tamilnadu, India. [2]National Center for Biological Sciences, Bangalore, India. [3]Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan.

Authors' contributions

MMG, RS and KF conceived the project. MMG carried out the computations and wrote the manuscript. KH modeled the structures of ORs and analyzed them. MMG, KH, RS and FK contributed in discussions. KH and RS wrote the sections about modeling. All authors read and finalized the manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Zhao H, Ivic L, Otaki JM, Hashimoto M, Mikoshiba K, Firestein S: **Functional expression of a mammalian odorant receptor.** *Science* 1998, **279**:237-242.
2. Kato A, Katada S, Touhara K: **Amino acids involved in conformational dynamics and G protein coupling of an odorant receptor: targeting gain-of-function mutation.** *J Neurochem* 2008, **107**:1261-1270.
3. Buck LB: **Information coding in the vertebrate olfactory system.** *Annu Rev Neurosci* 1996, **19**:517-544.
4. Mombaerts P: **Genes and ligands for odorant, vomeronasal and taste receptors.** *Nat Rev Neurosci* 2004, **5**:263-278.
5. Leal WS: **Behavioural neurobiology: The treacherous scent of a human.** *Nature* 2010, **464**:37-38.
6. Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR: **Odorant reception in the malaria mosquito Anopheles gambiae.** *Nature* 2010, **464**:66-71.
7. Jaén C, Ozdener MH, Reisert J: **Mechanisms of chloride uptake in frog olfactory receptor neurons.** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2011, **197**:339-349.
8. Hall RA: **Autonomic modulation of olfactory signaling.** *Sci Signal* 2011, **4**: pe1.
9. Abuin L, Bargeton B, Ulbrich MH, Isacoff EY, Kellenberger S, Benton R: **Functional architecture of olfactory ionotropic glutamate receptors.** *Neuron* 2011, **69**:44-60.
10. Ben-Chaim Y, Cheng MM, Yau KW: **Unitary response of mouse olfactory receptor neurons.** *Proc Natl Acad Sci USA* 2011, **108**:822-827.
11. Baud O, Etter S, Spreafico M, Bordoli L, Schwede T, Vogel H, Pick H: **The mouse eugenol odorant receptor: structural and functional plasticity of a broadly tuned odorant binding pocket.** *Biochemistry* 2011, **50**:843-853.
12. Ezkurdia L, Bartoli L, Fariselli P, Casadio R, Valencia A, Tress ML: **Progress and challenges in predicting protein-protein interaction sites.** *Brief Bioinform* 2009, **10**:233-246.
13. Yanay O, Rost B: **Protein-protein interaction hotspots carved into sequences.** *PLoS Comput Biol* 2007, **3**:e119.
14. You ZH, Lei YK, Huang DS, Zhou X: **Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data.** *Bioinformatics* 2010, **26(21)**:2744-2751.
15. Xia JF, Zhao XM, Song J, Huang DS: **APIS: accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility.** *BMC Bioinformatics* 2010, **11**:174.
16. Xia JF, Zhao XM, Huang DS: **Predicting protein-protein interactions from protein sequences using meta predictor.** *Amino Acids* 2010, **39**:1595-1599.
17. Gromiha MM, Yokota K, Fukui K: **Energy based approach for understanding the recognition mechanism in protein-protein complexes.** *Mol Biosyst* 2009, **5(12)**:1779-1786.
18. Gromiha MM, Yabuki Y, Suresh MX, Thangakani AM, Suwa M, Fukui K: **TMFunction: database for functional residues in membrane proteins.** *Nucleic Acids Res* 2009, **37**:D201-D204.
19. Kuang D, Yao Y, Wang M, Pattabiraman N, Kotra LP, Hampson DR: **Molecular similarities in the ligand binding pockets of an odorant receptor and the metabotropic glutamate receptors.** *J Biol Chem* 2003, **278**:42551-42559.
20. Luu P, Acher F, Bertrand HO, Fan J, Ngai J: **Molecular determinants of ligand selectivity in a vertebrate odorant receptor.** *J Neurosci* 2004, **24**:10128-10137.
21. Katada S, Hirokawa T, Oka Y, Suwa M, Touhara K: **Structural basis for a broad but selective ligand spectrum of a mouse olfactory receptor: mapping the odorant-binding site.** *J Neurosci* 2005, **25**:1806-1815.
22. Schmiedeberg K, Shirokova E, Weber HP, Schilling B, Meyerhof W, Krautwurst D: **Structural determinants of odorant recognition by the human olfactory receptors OR1A1 and OR1A2.** *J Struct Biol* 2007, **159**:400-412.
23. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M: **Crystal**

structure of rhodopsin: a G protein-coupled receptor. *Science* 2000, **289**:739-745.

24. Man O, Gilad Y, Lancet D: **Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons.** *Protein Sci* 2004, **13**:240-254.

25. Olender T, Feldmesser E, Atarot T, Eisenstein M, Lancet D: **The olfactory receptor universe - from whole genome analysis to structure and evolution.** *Genet Mol Res* 2004, **3**:545-553.

26. Tomii K, Kanehisa M: **Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins.** *Protein Eng* 1996, **9**:27-36.

27. Gromiha MM, Oobatake M, Sarai A: **Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins.** *Biophys Chem* 1999, **82**:51-67.

28. Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A: **Importance of surrounding residues for protein stability of partially buried mutations.** *J Biomol Struct Dyn* 2000, **18**:281-295.

29. Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A: **Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations.** *Protein Eng* 1999, **12**:549-555.

30. Gromiha MM, Selvaraj S: **Important amino acid properties for determining the transition state structures of two-state protein mutants.** *FEBS Lett* 2002, **526**:129-134.

31. Huang LT, Gromiha MM: **Reliable prediction of protein thermostability change upon double mutation from amino acid sequence.** *Bioinformatics* 2009, **25**:2181-2187.

32. Huang LT, Gromiha MM: **First insight into the prediction of protein folding rate change upon point mutation.** *Bioinformatics* 2010, **26**:2121-2127.

33. Ou YY, Chen SA, Gromiha MM: **Classification of transporters using efficient radial basis function networks with position-specific scoring matrices and biochemical properties.** *Proteins* 2010, **78**:1789-1797.

34. Tusndy GE, Simon I: **The HMMTOP transmembrane topology prediction server.** *Bioinformatics* 2001, **17**:849-850.

35. Pirovano W, Feenstra KA, Heringa J: **PRALINETM: a strategy for improved multiple alignment of transmembrane proteins.** *Bioinformatics* 2008, **24(2)**:492-497.

36. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ: **Jalview version 2–a multiple sequence alignment editor and analysis workbench.** *Bioinformatics* 2009, **25**:1189-1191, doi: 10.1093/bioinformatics/btp033.

37. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779-815.

38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **PROCHECK: a program to check the stereochemical quality of protein structures.** *J Appl Crystallogr* 1993, **26**:283-291.

39. Gromiha MM, Selvaraj S: **Inter-residue interactions in protein folding and stability.** *Prog Biophys Mol Biol* 2004, **86**:235-277.

40. Gromiha MM: **Influence of long-range contacts and surrounding residues on the transition state structures of proteins.** *Anal Biochem* 2011, **408**:32-36.

41. Grewal PS: *Numerical Methods of Statistical Analysis* New Delhi: Sterling Publ; 1987.

42. Witten IH, Frank E: *Data Mining: Practical Machine Learning Tools and Techniques.* 2 edition. San Francisco: Morgan Kaufmann; 2005.

43. Gromiha MM, Suwa M: **Discrimination of outer membrane proteins using machine learning algorithms.** *Proteins* 2006, **63**:1031-1037.

44. Gromiha MM, Yabuki Y: **Functional discrimination of membrane proteins using machine learning techniques.** *BMC Bioinformatics* 2008, **9**:135.

45. Gromiha MM, Suwa M: **Influence of amino acid properties for discriminating outer membrane proteins at better accuracy.** *Biochim Biophys Acta* 2006, **1764**:1493-1497.

46. Gromiha MM, Sowdhamini R, Fukui K: **Structure-function relationship in olfactory receptors.** *Lect Notes Bioinf* 2011, **6840**:618-623.

47. Lee B, Richards FM: **The interpretation of protein structures: estimation of static accessibility.** *J Mol Biol* 1971, **55(3)**:379-400.

48. Overington J, Johnson MS, Sali A, Blundell TL: **Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction.** *Proc Biol Sci* 1990, **241**:132-145.

49. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP: **JOY: protein sequence-structure representation and analysis.** *Bioinformatics* 1998, **14**:617-623.