PROTEINS: Structure, Function, and Bioinformatics 63:1018–1030 (2006)

A Two-Entropies Analysis to Identify Functional Positions in the Transmembrane Region of Class A G Protein-Coupled Receptors

Kai Ye,[1,2] Eric-Wubbo M. Lameijer,[1] Margot W. Beukers,[1] and Adriaan P. IJzerman[1]*
[1]*Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research, Leiden University, Leiden, The Netherlands*
[2]*College of Pharmacy, Wuhan University, Wuhan, China*

ABSTRACT **Residues in the transmembrane region of G protein-coupled receptors (GPCRs) are important for ligand binding and activation, but the function of individual positions is poorly understood. Using a sequence alignment of class A GPCRs (grouped in subfamilies), we propose a so-called "two-entropies analysis" to determine the potential role of individual positions in the transmembrane region of class A GPCRs. In our approach, such positions appear scattered, while largely clustered according to their biological function. Our method appears superior when compared to other bioinformatics approaches, such as the evolutionary trace method, entropy–variability plot, and correlated mutation analysis, both qualitatively and quantitatively. Proteins 2006;63:1018–1030.**

Key words: two-entropies analysis; G protein-coupled receptors; transmembrane region; evolutionary trace; correlated mutation analysis; entropy–variability plot; ligand-binding site

INTRODUCTION

G protein-coupled receptors (GPCRs) are integral cell membrane proteins that play a crucial role in signal transduction.[1–4] After binding of an endogenous ligand, such as a biogenic amine, peptide, nucleotide, or even protein, GPCRs undergo a conformational change leading to the activation of heterotrimeric G proteins. GPCRs are very successful drug targets since 30–45% of current drugs interact with this class of proteins.[5,6] Consequently, GPCRs represent up to 30% of the portfolio of many pharmaceutical companies.[7]

Despite numerous sequence–function studies on a large number of GPCRs (mainly on class A GPCRs, which represent more than 80% of all GPCRs according to the GPCR Data Base [GPCRDB][8]), at least two fundamental questions remain: (1) Which residues are responsible for the activation mechanism? (2) Which residues are critical for endogenous ligand binding? Because GPCRs comprise one of the largest superfamilies in the human genome, with almost 1000 proteins,[9] various bioinformatics approaches based on multiple sequence alignment have shed light on these two questions by identifying functional positions, especially at the binding site.[10–18] For example, using sequence pattern discovery techniques, Attwood created a database of hierarchical GPCR sequence fingerprints, from superfamily through family to receptor subtype levels.[10,12,13,15] The fingerprints identified at the family level show a certain correlation to the endogenous ligand binding, whereas the evolutionary trace method has recently been used to reveal global and subfamily-specific conserved residues of class A GPCRs.[17] It was reported that globally conserved residues relate to a canonical conformational switch, while some class-specific conserved residues form part of the ligand binding pocket.[17] Correlated mutation analysis and entropy–variability plots were also used to detect networks of functional residues in GPCRs.[11,14] The combination of these latter two methods allowed the identification of three groups of positions: residues responsible for G protein coupling, residues in the binding site, and residues in between these two groups.[11,14] However, the same studies revealed that the resolution (i.e., the capability of unambiguously assigning position to function) is often only modest.

In this study, we developed a new method, called two-entropies analysis, to identify functional positions of class A GPCRs. The multiple sequence alignment of class A GPCRs was divided into 70 subfamilies based on recognition of identical endogenous ligands.[8,19–21] Then we clustered functional positions based on two observations. The first observation is the variability of each position of class A GPCRs as a whole. The second one is the overall variability of each position within every individual subfamily. We reasoned that positions in a ligand-binding site are conserved within subfamilies but divergent among subfamilies. Positions that participate in folding/activation will be largely conserved both within subfamilies and among all class A GPCRs. This principle of predicting positions has been proposed and tried recently using different algo-

The Supplementary Materials referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585.suppmat

*Correspondence to: Adriaan P. IJzerman, Division of Medicinal Chemistry, Leiden/Amsterdam Center for Drug Research, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands. E-mail: ijzerman@lacdr.leidenuniv.nl

Received 1 July 2005; Revised 20 October 2005; Accepted 10 November 2005

Published online 10 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20899

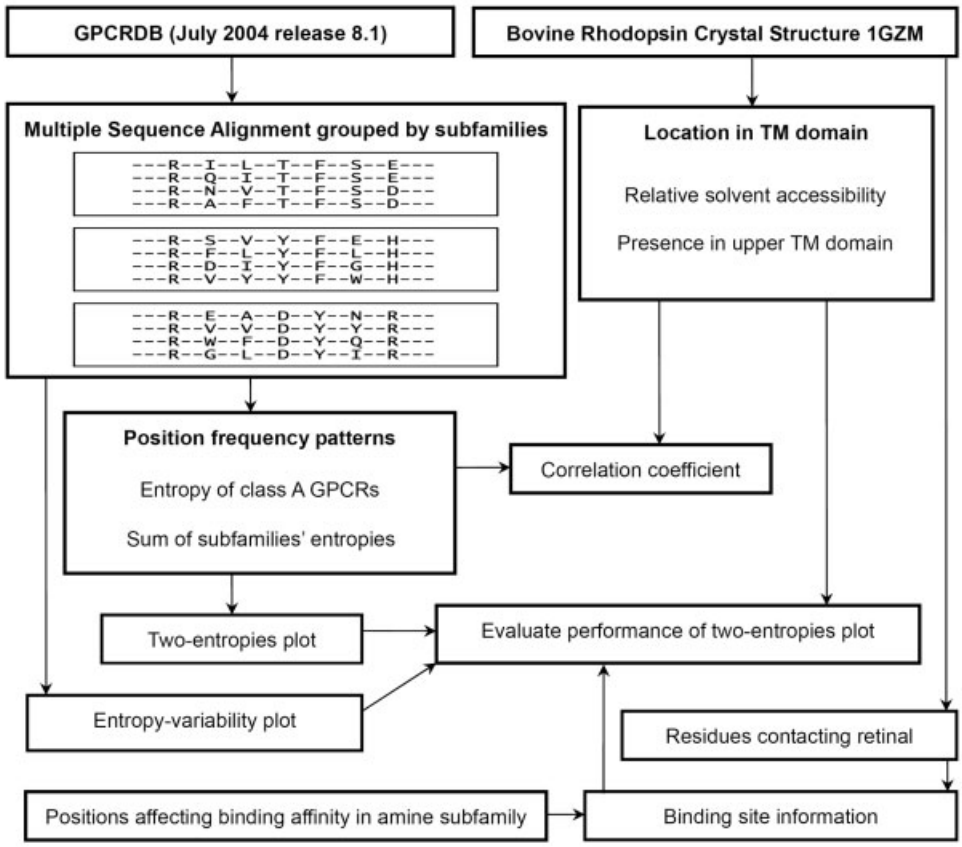

Fig. 1. Schematic flowchart of the method used to create and evaluate our two-entropies analysis for class A GPCRs.

rithms on protein families such as protein kinases, bacterial transcription factors, and olfactory receptors.[18,22–25] However, none of these methods have been applied to the whole set of class A GPCRs.

By comparing our predictions with structural and experimental data, as well as other bioinformatics approaches applied to GPCRs,[10–18] we were able to provide a global overview of functional positions in the transmembrane (TM) region of class A GPCRs. Moreover, our two-entropies analysis proved to be more discriminative than other methods.

MATERIALS AND METHODS

Our approach to cluster positions of class A GPCRs according to their functions and to evaluate the clustering results is depicted in Figure 1 and explained in detail as follows.

Sequences and Sequence Alignments

All subfamilies of class A GPCRs in the receptor compendium issued by the International Union of Pharmacology (IUPHAR)[21] were examined in this study. The sequence alignment of 1935 class A GPCRs that belong to these subfamilies were extracted from the GPCRDB (July 2004 release 8.1; http://www.gpcr.org/7tm/).[8] Then according to IUPHAR,[21] the 1935 Class A GPCRs were divided into 70 subfamilies based on recognition of identical endogenous ligands.

Numbering Scheme

To facilitate a consistent comparison of aligned residues in different class A GPCRs, we used the indexing method introduced by Ballesteros and Weinstein,[26] in which the most conserved residue in each TM helix is given the index number 50. The other residues are numbered relative to this position.

Definition of Boundaries of Seven Transmembrane Helix Regions

The definition of the start and end of the seven TM helices was adapted from the GPCRDB.[8] In the numbering scheme of Ballesteros and Weinstein,[26] these boundaries are as follows: 1.31 to 1.55 for TM1; 2.42 to 2.66 for TM2; 3.24 to 3.55 for TM3; 4.41 to 4.62 for TM4; 5.38 to 5.58 for TM5; 6.32 to 6.56 for TM6; and 7.29 to 7.52 for TM7.

Two Entropies Measures

To discriminate amino acid positions that participate in various functions such as binding and folding/activation, we defined two conservation measures based on the multiple sequence alignment of the entire class A GPCRs and the 70 subfamilies.

Entropy is defined here as a measure of conservation of amino acid residues at a certain position in a defined multiple sequence alignment. The relative frequency F_{ia} of residue type a at alignment position i in a given multiple sequence alignment with m proteins is given by

$$F_{ia} = Number_{ia}/m ,$$

where $Number_{ia}$ is the number of proteins that have residue type a at alignment position i.

The Shannon entropy at position i in the given alignment is given by the equation

$$E_i = -\sum_{a=1}^{20} F_{ia} \ln F_{ia} ,$$

where a loops over the 20 natural amino acids.

Two types of entropies were calculated. We calculated the first entropy, the entropy of the entire class A GPCRs, by taking the alignment of all class A GPCRs as input. The second entropy we calculated as the sum of the entropies of all subfamilies of class A GPCRs, while the entropy of each subfamily we calculated by taking the individual subfamily sequence alignment as input.

The smaller the entropy value, the more conserved the position is in the given sequence alignment.

Presence in Upper or Lower Domain of the Transmembrane Region

The positions within the TM region of class A GPCRs were divided into two subsets according to their presence in the upper or lower domain of the TM region. The definition of these two domains was adapted from the method of Imai and Fujita.[27] If a position is in the upper domain, a score of 1 is assigned. If a position is in the lower domain, a score of 0 is assigned.

Relative Solvent Accessibility

The solvent-accessible surface area of an amino acid residue indicates its level of burial (or solvent exposure) in a protein structure and is often expressed in terms of relative solvent accessibility (RSA). The RSA of a position i of a class A GPCR(RSA_i) was calculated using the template 1GZM,[28] the crystal structure of bovine rhodopsin. It is defined as the ratio of the solvent-exposed surface area of a residue X in position i of the bovine rhodopsin crystal structure, denoted as SA_i, and the maximum value of the solvent-exposed surface area for this amino acid corresponding to the surface-exposed area of the central residue observed in the tripeptide GXG in extended conformation, denoted as MSA_i. Thus, RSA_i adopts values between 0% and 100%, with 0% corresponding to a fully buried residue, and 100% to a fully accessible residue. The computer program MOLMOL 2K.2[29] was used to compute SA_i, MSA_i, and RSA_i for positions in the TM region of the crystal structure of bovine rhodopsin 1GZM.[28]

A two-state description distinguishing between residues that are buried (relative solvent accessibility $< 15\%$) and exposed (relative solvent accessibility $> 15\%$) was used.[30]

Correlation Matrices of Measures

We considered the two forms of entropy described, the RSA, and presence in the upper domain of the TM region as measures for the positions' properties. In order to provide a similarity score between these measures, Pearson correlations were performed for each of the two measures to create correlation matrices, which indicate the distance between every two measures. The Pearson correlation coefficient between any two series of numbers $x = \{x_1,x_2,\ldots,x_n\}$ and $y = \{y_1,y_2,\ldots,y_n\}$ is defined as

$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{X}}{\sigma_x}\right)\left(\frac{Y_i - \bar{Y}}{\sigma_y}\right) ,$$

in which $\bar{X}$ and $\bar{Y}$ are the average of these two series of numbers; σ is the standard deviation of these two series of numbers:

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{Y})^2} ,$$

The Pearson correlation coefficient is always between -1 and 1, with 1 meaning that two series are positively correlated, 0 meaning that they are completely uncorrelated, and -1 meaning they are perfectly negatively correlated.

Binding Site of Bovine Rhodopsin Based on Crystal Structure

To define the binding site of bovine rhodopsin, we used the crystal structure 1GZM.[28] Residues within 4 Å distance of the endogenous ligand retinal were considered to be part of the ligand-binding site. Calculation was performed with Deepview.[31]

Binding Sites of Aminergic Receptors Based on Experimental Data

Information about the binding site of aminergic receptors was derived from Shi and Javitch.[32] Positions were considered to be part of the ligand-binding site when they were located in the TM region and implicated in ligand binding in aminergic receptors based on experiments that address affinity labeling, functional complementation of mutations with modifications of ligand, or changes in antagonist affinity.

Entropy–Variability Plots of Class A GPCRs

The graphical representation of our two-entropies analysis is similar to an entropy–variability plot. To evaluate the performance of a two-entropies plot in separating positions with different functions, the entropy–variability plots of class A GPCRs were reproduced according to the method of Oliveira et al.[11] and served as a control.

Receiver–Operator Characteristic (ROC) Graph

Receiver–operator characteristic (ROC) graphs provide a visual tool for examining prediction performance.[33,34] An ROC graph is a plot with the false-positive rate on the x axis and the true-positive rate on the y axis. It is independent of class distribution or error costs.[34]

We made two ROC graphs to visualize the quantitative comparison of our two-entropies analysis with previous bioinformatics methods in predicting the ligand-binding site.

RESULTS

Correlation Coefficient Between Measures

The two types of entropy, the RSA, and the residue's presence in the upper domain of the TM region were used as measures. The correlation coefficients between these measures were calculated and are summarized in Table I.

TABLE I. Correlation Coefficient Between Measures

	Entropy of class A GPCRs	Presence in the upper domain	Relative solvent accessibility
Sum of subfamilies' entropies	0.678**	0.168*	0.665**
Entropy of class A GPCRs	—	0.401**	0.198**
Presence in upper domain	—	—	−0.042

**Correlation is significant at the .01 level ($P < .01$).
*Correlation is significant at the .05 level ($P < .05$).

Some measures were highly correlated, for instance, the two types of entropy. This was to be expected, because the positions conserved in all class A GPCRs will obviously be conserved in subfamilies too, and those that are divergent in proteins within a subfamily will be divergent in the entire class A GPCRs.

Separating Positions of the Upper and Lower Domain of the Transmembrane Region

Other correlations support what is known about the sequence–structure relationship in GPCRs, as reviewed in the Introduction. For example, the correlation coefficient between entropy of class A GPCRs and the presence of residues in the upper domain of the TM region is .401. This correlation coefficient means that the positions in the lower domain are significantly more conserved than those in the upper domain. For the positions in the upper domain (score of presence in the upper domain is 1), the entropy values of class A GPCRs are generally larger; for the positions in the lower domain (score of presence in the lower domain is 0), the entropy values of class A GPCRs are largely smaller. The positions in the upper domain involved in ligand binding appear to form a subfamily-specific binding site. As for the positions in the lower domain, they are conserved to maintain a similar overall fold and to evoke a similar cascade of activation events.

In Figure 2, we compared the performance of the two-entropies plot versus the entropy–variability plots in separating positions with respect to upper domain (red dots) and lower domain (blue triangles) in the TM region. Both methods illuminate the separation of the two domains.

Separating Positions With Different Relative Solvent Accessibility

The correlation between RSA and the sum of subfamilies' entropies of all class A GPCRs was more significant than the one between RSA and the entropy of class A GPCRs as a whole (Table I). Apparently, the positions on the surface of the receptors are more divergent than those in the core [blue triangles in Fig. 3(a)]. Most positions with high solvent accessibility have higher entropy values for the entire class A GPCRs. However, dozens of positions in the upper left corner of the two-entropies plot with large entropy values for the entire class A GPCRs (y axis) and a small sum of the subfamilies' entropies (x axis) have low solvent accessibility. This suggests that although these positions are in the core of receptors, they are divergent among class A GPCRs but conserved within subfamilies.

The performance of the two-entropies plot [Fig. 3(a)] in separating positions in the core from those on the surface of the TM regions was evaluated and compared with the entropy–variability plot of class A GPCRs [Fig. 3(b)]. Although we used the same entropy of class A GPCRs as a measure on the y axis, the sum of subfamilies' entropies (x axis) performed better than variability in not only providing a more distinct separation of positions with high variability but also in grouping positions with a similar level of burial in the receptor.

Separating Positions in the Ligand-Binding Site From Other Positions in the Transmembrane Region

We collected information about the ligand-binding site from both structural and biological data, and evaluated the performance of the two-entropies plot in separating positions at the binding site from other positions in the TM region. The binding site of bovine rhodopsin was taken from the crystal structure 1GZM, then mapped onto both the two-entropies plot and the entropy–variability plot, as shown in Figure 4. The residues within 4 Å distance of retinal in the crystal structure 1GZM are as follows: E113 (3.28), A117 (3.32), T118 (3.33), G121 (3.36), E122 (3.37), M207 (5.42), F212 (5.47), F261 (6.44), W265 (6.48), Y268 (6.51), A292 (7.39), and K296 (7.43).

Most positions contacting the ligand of bovine rhodopsin are indeed conserved within subfamilies but show great diversity among different subfamilies [upper left corner of Figure 4(a)]. However, a few positions, such as 5.47, 6.44, 6.48, and 6.51, are conserved with small entropy values with respect to both entropy of class A GPCRs and sum of subfamilies' entropies [lower left corner of Fig. 4(a)]. Those conserved positions that contact retinal are exclusively aromatic residues: F212 (5.47), F261 (6.44), W265 (6.48), and Y268 (6.51), which have been considered as an "aro-

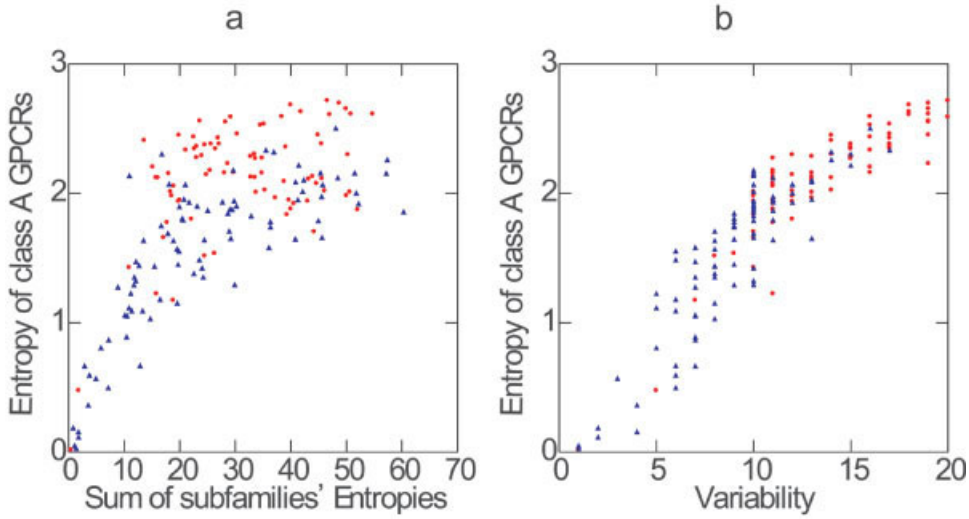

Figure 2.

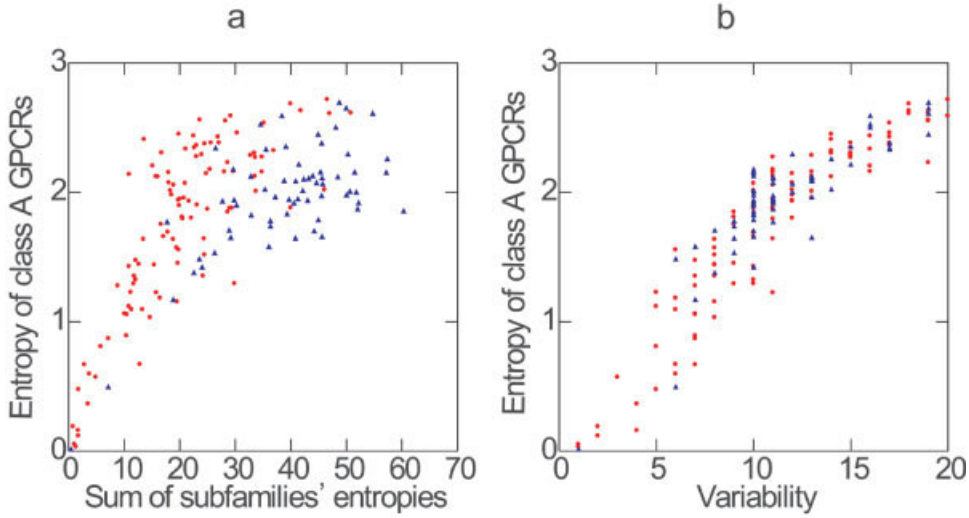

Figure 3.

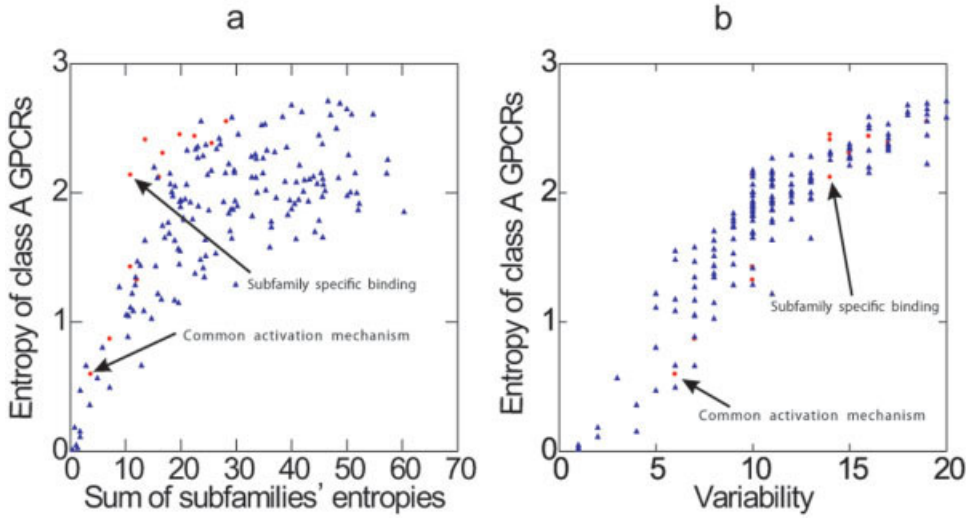

Figure 4.

matic cluster" before by Visiers et al.[35] According to the authors, once the ligand is recognized by subfamily-specific residues and occupies the binding region, the aromatic cluster will be disturbed and respond through concerted conformational rearrangements of the aromatic side-chains to promote receptor activation toward the cytoplasmic side of the receptor. It is safe to conclude that this conserved "aromatic cluster" makes no contribution to the specificity of endogenous ligand binding, but that it is responsible for a general activation mechanism [arrow pointing to "common activation mechanism" in Fig. 4(a)].

Thus, in the two-entropies plot [Fig. 4(a)], a cluster of positions shows up at the upper left corner, where the sum of subfamilies's entropies is small and entropy of class A GPCRs is large. These positions probably represent the ligand-binding site (arrow pointing to "subfamily-specific binding"). However, these positions that may determine subfamily-specific binding are mixed with other positions in the entropy–variability plot [Fig. 4(b)].

Positions within the TM region implicated in ligand binding in aminergic receptors based on affinity labeling, functional complementation of mutations with modifications of ligand, or changes in antagonist affinity[32] were mapped onto both the two-entropies plot and the entropy–variability plot. Although very often mutated residues that affect ligand binding are in the ligand-binding site, it is also possible that affinity changes are caused by indirect effects such as changing receptor folding or receptor surface expression level.[36] In this case, binding-site positions derived from the biological data would be distributed more widely in the two-entropies plot [Fig. 5(a)] compared to the binding site of bovine rhodopsin [Fig. 4(a)]. However, most of these positions are still clustered at the upper left corner of the two-entropies plot, where we suggest the subfamily-specific binding region to be [Fig. 5(a)].

Similarly, as in Figure 4(b), the entropy–variability plot [Fig. 5(b)] did not provide a good separation between binding sites and other positions.

Fig. 2. Comparison of the performance of the two-entropies plot and the entropy–variability plot in separating positions in the upper and lower domains of the TM region of class A GPCRs. Red dots are positions in upper domain; blue triangles are positions in lower domain. (**a**) The *x* axis is sum of subfamilies' entropies for each position in the TM region; the *y* axis is entropy of class A GPCRs for each position in the TM region. (**b**) The *x* axis is variability for each position in the transmembrane region; the *y* axis is entropy of class A GPCRs for each position in the transmembrane region.

Fig. 3. Comparison of the performance of our two-entropies plot and the entropy–variability plot in separating positions with different RSA. Red dots are positions with relative solvent accessibility smaller than 15%; blue triangles are positions with RSA larger than 15%. For labeling of axes, see Figure 2.

Fig. 4. Comparison of the performance of the two-entropies plot and the entropy–variability plot in separating ligand-binding site positions from other positions using information derived from the bovine rhodopsin crystal structure 1GZM. Red dots are positions within 4 Å distance to retinal, the ligand of bovine rhodopsin; blue triangles are positions with greater than 4 Å distance to retinal. For labeling of axes, see Figure 2.

Clusters of Positions Identified by the Two-Entropies Plot

According to Figures 2–5, positions in the TM region of class A GPCRs in our two-entropies plot tend to cluster according to their functions. After manually mapping positions onto the crystal structure of bovine rhodopsin, we suggest dividing these positions into six clusters (Fig. 6). The positions in cluster 1 are those that frequently participate in endogenous ligand binding, such as position 3.32. These positions in cluster 1 are in the upper domain of the TM region, as well as in the core of receptors. They are conserved within subfamilies but divergent among subfamilies. The positions in cluster 2 are involved in folding, such as C3.25, which forms a disulfide bridge with a cysteine residue in extracellular loop 2, or in activation, such as W6.48 and R3.50. Most positions in cluster 2 are in the lower domain of the TM region and also in the core of receptors. They are conserved among all class A GPCRs. The positions in cluster 3 are at the extracellular end of helices. Among those 21 positions in cluster 3, four positions are at the extracellular end of helix 1, three positions are at the extracellular end of helix 2, and seven positions are at the extracellular end of helix 7. The positions in cluster 3 are not conserved within and among subfamilies. However, these positions are clustered together in space, and they are not far away from the potential binding site. This finding may have potential in drug research, in that synthetic ligands may be modified to contact those positions to achieve better receptor subtype selectivity. The positions in cluster 4 are slightly less conserved than the 16 positions of cluster 1 with respect to being either within or among subfamilies. They are located primarily in the lower domain of the TM region and in the core of receptors, and are probably involved in helix–helix interaction to conserve the receptor's architecture and to provide a similar activation mechanism for class A GPCRs. Mutation of the positions in cluster 4 can cause receptor constitutive activity, for example, positions 3.43 and 6.37.[37–44] The positions in cluster 5 mostly face the cell membrane. They are divergent both within subfamilies and among all class A GPCRs. However, they are less divergent than positions in cluster 1 with respect to the entire class A GPCRs. Presumably, amino acids with various properties will occur in the ligand binding site (cluster 1) of receptors to accommodate variation of ligands in shape, electrostatic and H-bond interactions, and aromatic stacking, for example, position 3.32 (charged KRHDE, 29.63%; aromatic FYW, 17.84%; hydrophobic AVLI, 27.09%; polar but uncharged STCMNQ, 10.04%, G, 4.99%, P, 1.80%). However, for positions facing the membrane, hydrophobic amino acids are more dominant, for instance, position 4.47 (charged KRHDE, 0.85%; aromatic FYW, 2.02%; hydrophobic AVLI, 70.49%; polar but uncharged STCMNQ, 14.12%, G, 10.62%, P, 1.80%). The positions in cluster 6 are in the middle of clusters 1, 3, 4, and 5 such that the potential functions of these positions may be a mixture of functions of nearby clusters. For instance,

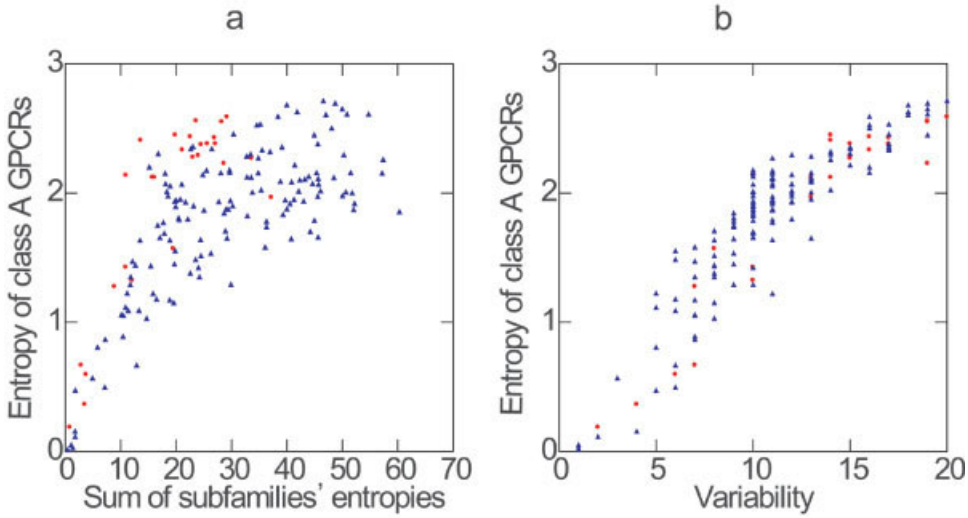

Fig. 5. Comparison of the performance of the two-entropies plot and entropy–variability plot in separating binding site positions from other positions. Red dots are positions within the TM region implicated in ligand binding in aminergic receptors based on experimental results; blue triangles are other positions in the TM region of class A GPCRs. For labeling of axes, see Figure 2.

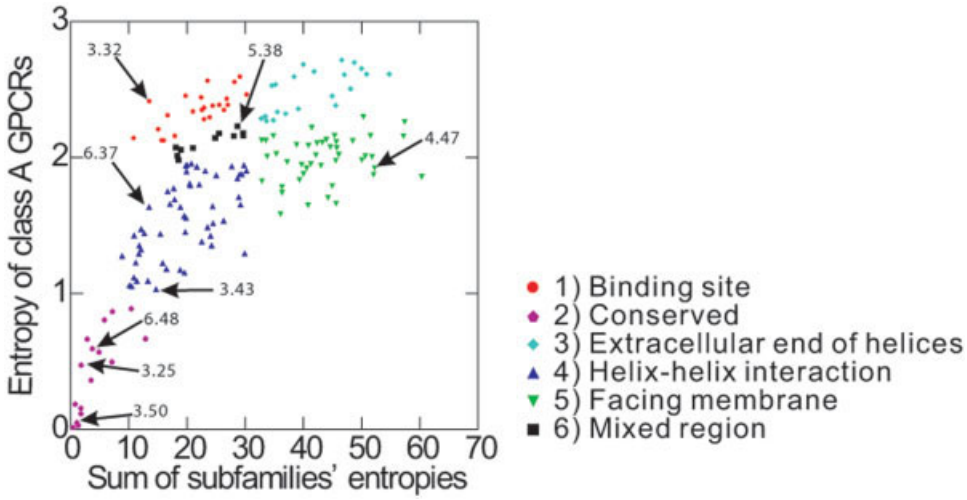

Fig. 6. Clustering positions in the two-entropies plot. For labelings of axes, see Figure 2. See Supplementary Material to get information about the positions labeled according to the numbering scheme of Ballesteros and Weinstein[26] and labeled according to the bovine rhodopsin sequence.

position 5.38, which is close to clusters 1, 3, and 5, is at the end of helix 5 and also at a feasible location to contact the ligand.

DISCUSSION

We divided the functions of positions in the TM region of class A GPCRs into three categories: binding, folding/activation, and "other." Previous studies have shown that strongly conserved positions such as C3.25, R3.50, and W6.48 (numbering scheme according to Ballesteros and Weinstein[26]) are involved in receptor folding and activation.[3,16,35,45–48] Our approach puts more emphasis on discriminating between the binding sites of class A GPCRs and the other two categories. It aims to cluster residues according to their function based on two assumptions. The first is that residues are largely conserved in the binding site of homologous receptors in the same subfamily binding the same endogenous ligand. If so, most GPCRs should share identical residues at binding sites if they belong to the same subfamily and bind an identical endogenous ligand. This is also the unaccounted assumption in both evolutionary trace and correlated mutation analysis during the process of identifying binding sites of GPCRs.[14,17] The second assumption is that endogenous binding sites are located in a region embedded between TM helices. This has been shown experimentally for a large number of receptors, including those for biogenic amines,[49] nucleotides,[50] melatonin,[51] and prostacyclin.[52]

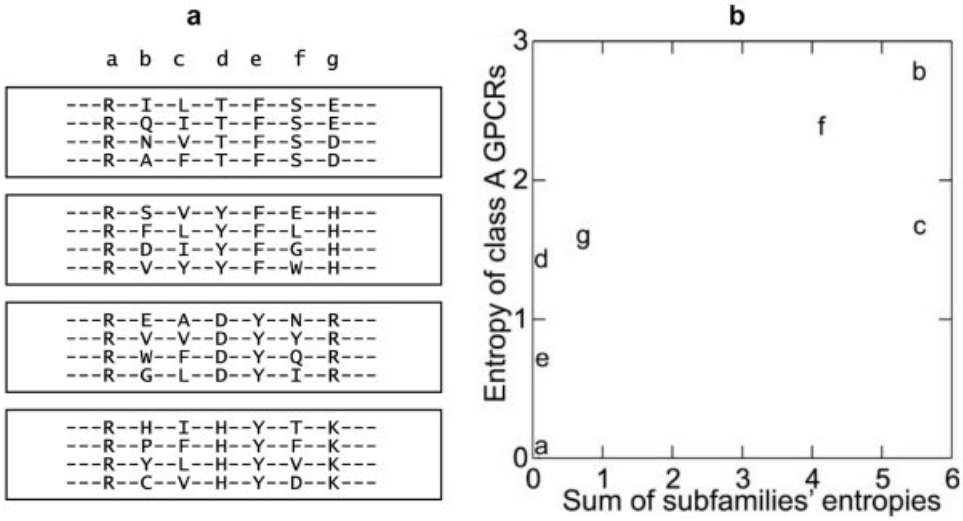

Fig. 7. (**a**) Pseudosequence alignment of four hypothetic subfamilies of class A GPCRs. Each subfamily has four fictitious sequence fragments. (**b**) Plotting of the positions in a two-entropies plot.

In the entropy–variability plot, both entropy and variability are measures of conservation of each position of class A GPCRs. The variability at a position is defined as the number of different residue types observed at this position in at least 0.5% of all sequences.[11,14] Thus, the entropy and the variability are strongly correlated, and positions in entropy–variability plots are crowded along the diagonal. For this reason, in our two-entropies plot, we only adopted one measure, entropy of all class A GPCRs, to separate overall conserved positions from divergent positions. The second measure, sum of subfamilies' entropies, was introduced to scatter positions with high entropy values of all class A GPCRs. In this way, our two-entropies plot achieves a better balance of robustness and sensitivity than the entropy–variability plot or the evolutionary trace method, which we discuss later.

Positions With Different Amino Acid Frequency Patterns Scattered in the Two-Entropies Plot

To compare the performance of our two-entropies plot with other sequence alignment–based methods in differentiating positions with different functions, we propose a sequence alignment for four hypothetical subfamilies of class A GPCRs. Each subfamily is suggested to bind a different endogenous ligand. The alignment was established within and also between subfamilies [Fig. 7(a)]. Please note that this hypothetical set of GPCR sequences was only used to illustrate the principle of our approach. As described in the Results section, our further analysis was based on the total set of 1935 GPCR sequences from 70 subfamilies.

The two entropies of each position in Figure 7(a) were calculated according to the algorithm described in the Methods section and are shown in Table II. The overall configuration of positions in Figure 7(b) is more pronounced than in Figures 2–6. This is due to a relatively small number of subfamilies [four subfamilies, Fig. 7(a)] and either perfect conservation or divergence in the sequence alignment [Fig. 7(a)]. For example, in a more realistic situation, position *d* indeed has a small sum of subfamilies' entropies, but not zero. Its entropy value of all subfamilies will be larger because many more subfamilies (70) are present in the alignment than in the four hypothetic ones.

TABLE II. Entropy of Sequence Alignment and Sum of Subfamilies' Entropies of Sequence Alignment in Figure 7(a)

	a	b	c	d	e	f	g
Entropy of all subfamilies	0	2.77	1.66	1.39	0.69	2.39	1.56
Sum of subfamilies' entropies	0	5.55	5.55	0	0	4.16	0.69

Our two-entropies plot [Fig. 7(b)] illustrates the differences between positions *a* through *e*. For position *a*, amino acids are conserved within and between subfamilies. Both entropies are small, and this position *a* will appear in the lower left corner of the two-entropies plot. Functions of position *a* could be folding, such as C3.25, which forms a disulfide bridge with a conserved cysteine residue in extracellular loop 2, or activation, such as W6.48 and R3.50.

As for position *b*, amino acids are neither conserved between subfamilies nor conserved in the individual subfamilies. In addition, hydrophilic, hydrophobic, and aromatic residues show up in position *b*. In this case, both entropy measures will have large values. The function of position *b* may be other than ligand binding and folding/activation.

Position *c* is quite similar to position *b*; amino acids are neither conserved between subfamilies nor conserved in each subfamily. However, only hydrophobic residues show up in position *c*. So both entropy measures have large

values, but smaller than those of position *b*. Residues in position *c* probably face the membrane.

Position *d* is very important. Although all 20 amino acids may show up at this position, they are mostly conserved within each subfamily but divergent among subfamilies. For position *d*, the sum of subfamilies' entropies will be small, and the entropy of class A GPCRs will be large. It is probable that position *d* frequently participates in endogenous ligand binding.

Position *e* is also very important. Residues are conserved within each subfamily and are also shared by several subfamilies. As a result, the sum of subfamilies' entropies will be small but the entropy of class A GPCRs will be larger than position *a* and smaller than positions *b*, *c*, and *d*. Position *e* may participate in helix–helix interactions to conserve the three-dimensional (3D) aspects of GPCRs and to provide a common activation mechanism for class A GPCRs.

Positions *f* and *g* are discussed later, to compare the evolutionary trace method with our two-entropies analysis.

Two-Entropies Plot Versus Entropy–Variability Plot

Although entropy–variability plots have been used very successfully in the past,[11,14,16] the performance of our two-entropies plot in separating positions according to their functions appears improved (Figs. 2–5). Most importantly, as shown in Figures 4 and 5, the entropy–variability plot does not differentiate very well between positions *b* and *d*. The explanation is as follows: When more than 20 subfamilies are present in one superfamily, 20 residue types are likely to be present in each position of the binding region to account for the diversity of endogenous ligands. This is the case for the large family of class A GPCRs; hence, the entropy value of position *b* will be as large as the one of position *d*, and the variability of both positions *b* and *d* will be close to 20.

Two-Entropies Plot Versus the Evolutionary Trace Method

The evolutionary trace method has been shown to be successful in predicting binding sites of soluble and membrane proteins.[53–59] Recently, this method was used to analyze class A GPCR sequences to identify globally conserved residues and opsin subfamily-specific residues.[17] In that study, only four subfamilies of class A GPCRs—visual opsin, bioamine, olfactory, and chemokine—were included to trace 39 "globally" conserved residues. Only the opsin subfamily was subjected to differential trace analysis, and finally 17 opsin "specific" conserved residues were identified.[17]

However, the identified 39 "globally" conserved residues based on only four subfamilies are not conserved in all subfamilies of class A GPCRs. For example, position 3.33 was identified as one of the 39 "globally" conserved residues. But great variation in position 3.33 is observed among all class A GPCRs (Fig. 8). Because it is hard to include dozens of subfamilies in the evolutionary trace method and to compare subfamily-specific conserved residues between every pair of subfamilies, we believe the evolutionary trace method does not make full use of the rich sequence information of a superfamily as large as class A GPCRs.

In addition, the evolutionary trace method does not easily recognize positions with small variation as globally conserved residues. For example, the method failed to identify position 2.50 as a globally conserved position,[17] which has D in 92% of class A GPCRs. Suppose that among 1000 proteins of class A GPCRs, only two amino acid types (e.g., D and K) are present in a certain position. Obviously, there is a great difference between a situation with 1D/999K versus 500D/500K. Unfortunately, the evolutionary trace method ignores such a difference and considers both of the above situations as a nonconserved position. However, both the entropy–variability plot and our two-entropies plot detect such a conservation, because they are designed to measure conservation on the basis of not only the number of amino acid types at a given position but also the frequency of each amino acid type at that position.

In principle, the evolutionary trace method differentiates between positions *a*, *b*, *d*, and *e* (Fig. 7). However, it may make mistakes at position *f*, which is conserved in just one subfamily, and hence considers position *f* as being in the ligand-binding site. Position *f* may not be functionally important, because it is possible that such "conservation" is caused by a small population of proteins or a short evolution history since the subfamily member began to evolve. In our two-entropies plot, position *f* is not misjudged, because our approach takes not just one subfamily into account but the overall observation in all subfamilies.

The evolutionary trace method may also lead to erroneous results in subfamilies in which ligand binding sites are not completely conserved, such as position *g* [Fig.7(a)]. For instance, in adenosine receptors, position 7.42 was reported to be involved in agonist binding.[60–64] However, position 7.42 is a serine in the human adenosine A_{2A} receptor but a threonine in the human adenosine A_1 receptor. Because of the high sensitivity to class-specific conservation, any slight variation at the binding site will impede the evolutionary trace method in identifying the binding site. In contrast, our two-entropies plot will still consider position *g* as belonging to the binding site, since the joint conservation within subfamilies and large divergence in all class A GPCRs strongly indicate that this position is involved in the ligand binding. For this reason, the evolutionary trace method probably failed to predict two positions, 3.32 and 3.37, as part of the binding site of bovine rhodopsin. However, these two positions are located at the upper left corner of the two-entropies plot, and they are indeed within 4 Å distance of retinal, the endogenous ligand of bovine rhodopsin.

Two-Entropies Plot Versus Sequence Pattern Discovery

Various sequence pattern discovery approaches have been applied to GPCRs. For example, using sequence pattern discovery techniques, Attwood created a database

of hierarchical GPCR sequence fingerprints, from superfamily through family to receptor subtype levels.[10,12,13,15] The fingerprints identified at the family level show a certain correlation to the endogenous ligand binding.

Compared to the sequence pattern discovery approaches, our approach predicts the functional sites of GPCRs in a more precise way for two reasons. First of all, our method exploits the conservation among all subfamilies rather than per subfamily. Second, out method can handle very large numbers of sequences at the same time. In contrast, sequence pattern discovery algorithms investigate only one phylogenetic level of subfamily each time, without a cross-check between different subfamilies. This defect does harm, as we can see in the evolutionary trace method discussed above: Some positions in the identified sequence patterns may be conserved in a certain subfamily, but they may not be functionally important, because such conservation is caused by the small population of the subfamily or the short evolution history since the subfamily member began to evolve. Thus, such approaches never take full advantage of the huge superfamily with dozens of subfamilies, such as GPCRs.

Two-Entropies Plot Versus Correlated Mutation Analysis

The method of correlated mutation analysis (CMA) was proposed to detect intramolecular or intermolecular contacts or links between residues. It successfully predicted the approximate binding region of class A GPCRs.[11,14] The principle of CMA in defining the binding region of class A GPCRs is that when one endogenous ligand changes to another, residues in the binding site will "mutate" at the same time to accommodate the change of ligand in shape, and in hydrophobic and electrostatic properties. CMA easily discriminates position *a* and position *e* in the pseudoalignment shown in Figure 7(a). As shown in Figure 9(a) , CMA is consistent with our two-entropies plot: Most binding sites of class A GPCRs predicted by CMA are clustered at the upper left corner of our two-entropies plot.

However, let us differentiate the residues close to the binding site into two layers. The first layer of residues surrounds the ligand, so that they are indeed at the binding site. The residues contacting the first layer of residues but not the ligand constitute the second layer of residues. When the first-layer residues "mutate" to accommodate a different ligand, the second-layer residues will also "mutate" at the same time to maintain compact contact and correct interaction with the first layer. The correlation between the first and the second layer may be so strong that it is hard to discriminate between the first layer and the second. For this reason, prediction of binding sites of class A GPCRs by CMA also includes positions facing the membrane, such as position 6.45 (Fig. 9). In addition, CMA failed to predict positions 5.39, 2.64, and 4.60 as involved in ligand binding (Fig. 9), mutation of which affected agonist binding affinity in aminergic receptors.[32] These three positions, however, are conserved within subfamilies and divergent between subfamilies, so that they are indeed part of the binding site of class A GPCRs according to the two-entropies plot. Finally, CMA will only detect a functional network where positions are either conserved or strongly correlated with other positions. For example, in Figure 9(a), many positions in the lower left corner of the two-entropies plot are "invisible" to CMA, because the frequencies of the 20 amino acids in these positions show no correlation with other positions, although these positions are functionally important.

Two-Entropies Plot Versus "Mutual Information"

Mirny and coworkers used a "mutual information" approach to measure both conservation within subfamily and diversity between subfamilies. The authors used various statistical models to evaluate significance of mutual information and to identify so-called "specificity-determining" positions.[22–24]

In our approach, we use two measures to evaluate the conservation of positions among proteins within the same subfamily and their diversity between different subfamilies. The combined two measures (*entropies*) result in a high resolution in identifying binding sites and other functional sites.

Although our method apparently provides a global overview of all functional positions, it is not known yet how well the two methods of two-entropies analysis and mutual information compare, since they have not been applied to the same data set. Thus, we are currently applying our method to the superfamily of protein kinases, which have already been analyzed by mutual information.[22]

Quantitative Comparison With Previous Bioinformatics Methods

Two ROC plots[34] were made to visualize the quantitative comparison of our two-entropies analysis with previous bioinformatics methods.

As mentioned above, the ligand-binding site (12 positions) of bovine rhodopsin was taken from the crystal structure 1GZM. We compared our two-entropy analysis and other bioinformatics methods using an ROC plot (Fig. 10). The binding-site information of aminergic receptors based on experimental data was also included in this comparison. Apparently, our two-entropies analysis performs better than other bioinformatics methods. If the binding-site information of aminergic receptors based on experimental data was used to predict the binding site of rhodopsin, it performed slightly better (correctly predicted 10 out of 12 amino acids) than our two-entropies analysis (predicted 9 of 12 residues). However, the three positions (F5.47, F6.44, and W6.48) that our method missed are conserved among class A GPCRs and belong to the highly conserved aromatic cluster, which does not contribute to subfamily-specific ligand binding, but instead is part of the activation cascade. In other words, our method successfully predicted all subfamily-specific residues of bovine rhodopsin that determine endogenous ligand binding and was able to discriminate between ligand-binding residues and activation residues.

The positions involved in rhodopsin ligand binding are only a subset of the general ligand-binding positions in

class A GPCRs as predicted by our two-entropies analysis. The ligand-binding residues that we have identified form a larger set of residues to accommodate the various sizes and shapes of the endogenous ligands of class A GPCRs. Thus, each endogenous ligand binds to a subset of these residues, and the binding site determined from the bovine rhodopsin crystal structure represents the residues that make up the retinal binding site. As a result, the binding site determined from the bovine rhodopsin crystal structure cannot well represent other class A GPCRs, as shown in the next ROC plot (Fig. 11).

As mentioned above, the information about the binding site of aminergic receptors was derived from Shi and Javitch.[32] When the binding-site information of bovine rhodopsin based on its crystal structure was used to predict the binding site of aminergic receptors, it performed much worse than our two-entropies analysis and CMA. Apparently, rhodopsin and aminergic receptors use different subsets of the general binding site as their ligand-binding site.

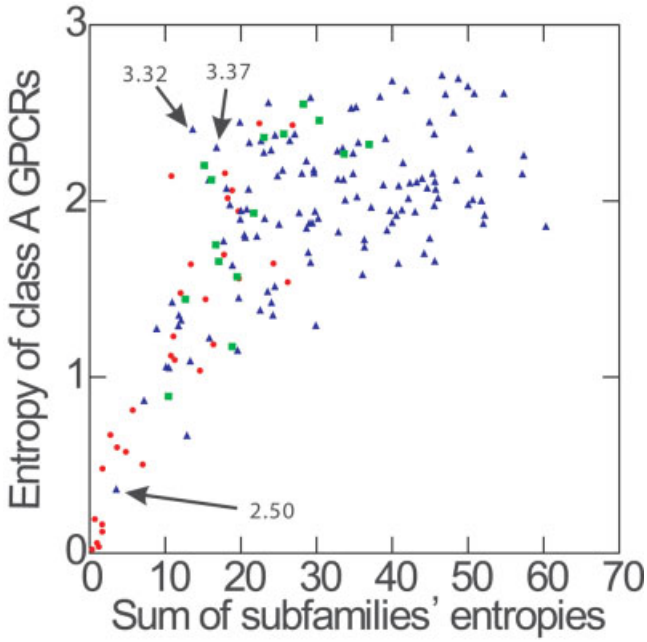

Fig. 8. Plotting "globally" conserved positions and opsin subfamily "specific" conserved positions[17] identified by the evolutionary trace method onto our two-entropies plot. The *x* axis is the sum of subfamilies' entropies; the *y* axis is entropy of class A GPCRs. Red dots are "globally" conserved positions; green squares are opsin subfamily "specific" conserved positions; blue triangles are other positions.

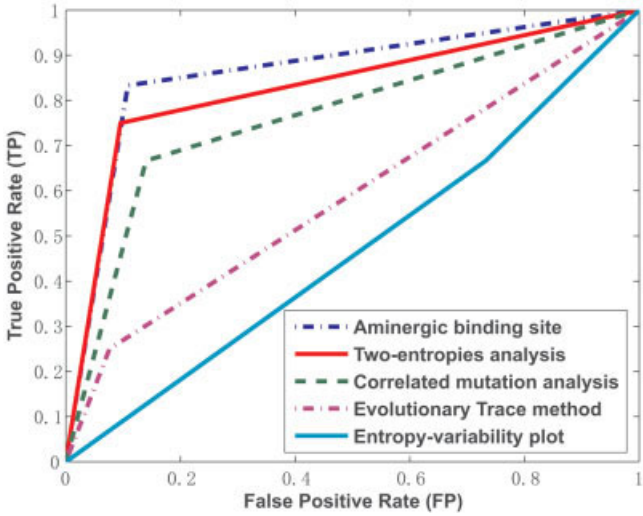

Fig. 10. ROC graph for the prediction of the ligand-binding site of bovine rhodopsin. True-positive and false-positive rates were calculated for each bioinformatics approach for the prediction of the ligand-binding site of bovine rhodopsin, using its crystal structure as "gold standard."

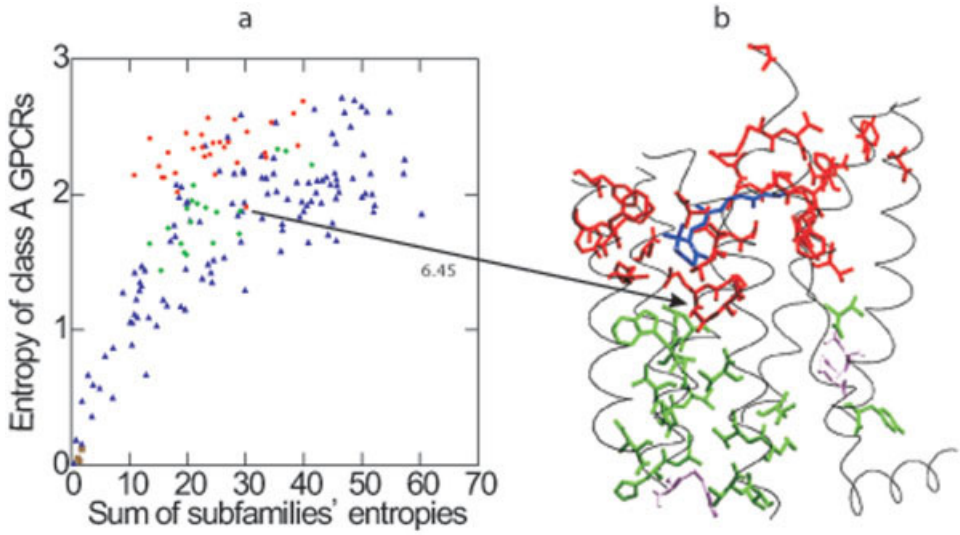

Fig. 9. Plotting three networks of positions identified by CMA[11] in the two-entropies plot (**a**) and mapping these three networks onto the crystal structure of bovine rhodopsin (**b**). Brown squares are conserved positions (network 1). Red dots are involved in ligand binding (network 2). Green rhombuses are involved in G protein coupling and activation (network 3).

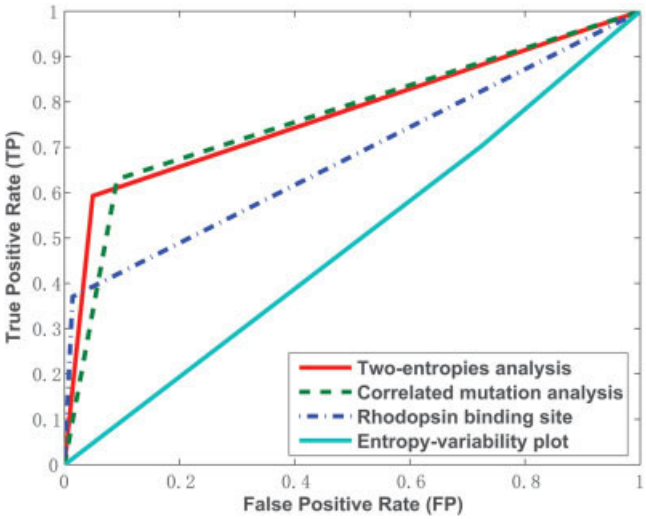

Fig. 11. ROC graph for the prediction of the ligand-binding site of aminergic receptors. True-positive and false-positive rates were calculated for each bioinformatics approach for the prediction of the ligand-binding site of aminergic receptors, using the experimental data as "gold standard."

CONCLUSION

Based on the sequence alignment of class A GPCRs grouped into subfamilies, a two-entropies analysis is proposed to determine the potential functions of positions in the TM region of GPCRs. In our two-entropies plot approach, positions of class A GPCRs in the TM region were scattered and clustered according to their biological functions. The two-entropies analysis may also be applicable to other protein superfamilies.

REFERENCES

1. Schoneberg T, Schulz A, Gudermann T. The structural basis of G-protein-coupled receptor function and dysfunction in human diseases. Rev Physiol Biochem Pharmacol 2002;144:143–227.
2. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. Nat Rev Mol Cell Biol 2002;3:639–650.
3. Gether U, Asmar F, Meinild AK, Rasmussen SG. Structural basis for activation of G-protein-coupled receptors. Pharmacol Toxicol 2002;91:304–312.
4. Gether U. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. Endocr Rev 2000;21:90–113.
5. Hopkins AL, Groom CR. The druggable genome. Nat Rev Drug Discov 2002;1:727–730.
6. Drews J. Drug discovery: a historical perspective. Science 2000;287: 1960–1964.
7. Klabunde T, Hessler G. Drug design strategies for targeting G-protein-coupled receptors. Chembiochem 2002;3:928–944.
8. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G. GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 2003;31:294–297.
9. Takeda S, Kadowaki S, Haga T, Takaesu H, Mitaku S. Identification of G protein-coupled receptor genes from the human genome sequence. FEBS Lett 2002;520:97–101.
10. Kuipers W, Oliveira L, Vriend G, IJzerman AP. Identification of class-determining residues in G protein-coupled receptors by sequence analysis. Receptors Channels 1997;5:159–174.
11. Oliveira L, Paiva AC, Vriend G. Correlated mutation analyses on very large sequence families. Chembiochem 2002;3:1010–1017.
12. Attwood TK, Blythe MJ, Flower DR, Gaulton A, Mabey JE, Maudling N, McGregor L, Mitchell AL, Moulton G, Paine K, Scordis P. PRINTS and PRINTS-S shed light on protein ancestry. Nucleic Acids Res 2002;30:239–241.
13. Attwood TK. The PRINTS database: a resource for identification of protein families. Brief Bioinform 2002;3:252–263.
14. Oliveira L, Paiva PB, Paiva AC, Vriend G. Identification of functionally conserved residues with the use of entropy–variability plots. Proteins 2003;52:544–552.
15. Attwood TK, Bradley P, Flower DR, Gaulton A, Maudling N, Mitchell AL, Moulton G, Nordle A, Paine K, Taylor P, Uddin A, Zygouri C. PRINTS and its automatic supplement, prePRINTS. Nucleic Acids Res 2003;31:400–402.
16. Oliveira L, Paiva PB, Paiva AC, Vriend G. Sequence analysis reveals how G protein-coupled receptors transduce the signal to the G protein. Proteins 2003;52:553–560.
17. Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG, Lichtarge O. Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. J Biol Chem 2004;279:8126–8132.
18. 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.
19. Horn F, Weare J, Beukers MW, Horsch S, Bairoch A, Chen W, Edvardsen O, Campagne F, Vriend G. GPCRDB: an information system for G protein-coupled receptors. Nucleic Acids Res 1998;26: 275–279.
20. Horn F, Vriend G, Cohen FE. Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. Nucleic Acids Res 2001;29:346–349.
21. Godfraind T, Vanhoutte P, Ruffolo R, Humphrey P. The IUPHAR compendium of receptor characterization and classification. London: IUPHAR Media; 1998. 267 p.
22. Li L, Shakhnovich EI, Mirny LA. Amino acids determining enzyme–substrate specificity in prokaryotic and eukaryotic protein kinases. Proc Natl Acad Sci USA 2003;100:4463–4468.
23. Mirny LA, Gelfand MS. Using orthologous and paralogous proteins to identify specificity determining residues. Genome Biol 2002;3:PREPRINT0002.
24. Mirny LA, Gelfand MS. Using orthologous and paralogous proteins to identify specificity-determining residues in bacterial transcription factors. J Mol Biol 2002;321:7–20.
25. Chiu HC, Chang CA, Hu YJ. Prediction of orthologous relationship by functionally important sites. Comput Methods Programs Biomed 2005;78:209–222.
26. Ballesteros JA, Weinstein H. Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G protein-coupled receptors. Methods Neurosci 1995;25:366–428.
27. Imai T, Fujita N. Statistical sequence analyses of G-protein-coupled receptors: structural and functional characteristics viewed with periodicities of entropy, hydrophobicity, and volume. Proteins 2004;56:650–660.
28. Li J, Edwards PC, Burghammer M, Villa C, Schertler GF. Structure of bovine rhodopsin in a trigonal crystal form. J Mol Biol 2004;343:1409–1438.
29. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55, 29–32.
30. Rost B, Sander C. Bridging the protein sequence–structure gap by structure predictions. Annu Rev Biophys Biomol Struct 1996;25: 113–136.
31. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
32. Shi L, Javitch JA. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. Annu Rev Pharmacol Toxicol 2002;42:437–467.
33. Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285–1293.
34. Provost F, Kohavi R. Guest editors' introduction: On applied research in machine learning. Machine Learning 1998;30:127–132.
35. Visiers I, Ballesteros JA, Weinstein H. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. Methods Enzymol 2002;343:329–371.
36. Kristiansen K, Kroeze WK, Willins DL, Gelber EI, Savage JE, Glennon RA, Roth BL. A highly conserved aspartic acid (Asp-155) anchors the terminal amine moiety of tryptamines and is involved in membrane targeting of the 5-HT(2A) serotonin receptor but

does not participate in activation via a "salt-bridge disruption" mechanism. J Pharmacol Exp Ther 2000;293:735–746.
37. Lu ZL, Curtis CA, Jones PG, Pavia J, Hulme EC. The role of the aspartate–arginine–tyrosine triad in the M1 muscarinic receptor: mutations of aspartate 122 and tyrosine 124 decrease receptor expression but do not abolish signaling. Mol Pharmacol 1997;51: 234–241.
38. Lu ZL, Hulme EC. The functional topography of transmembrane domain 3 of the M1 muscarinic acetylcholine receptor, revealed by scanning mutagenesis. J Biol Chem 1999;274:7309–7315.
39. Tao YX, Abell AN, Liu X, Nakamura K, Segaloff DL. Constitutive activation of G protein-coupled receptors as a result of selective substitution of a conserved leucine residue in transmembrane helix III. Mol Endocrinol 2000;14:1272–1282.
40. Min L, Ascoli M. Effect of activating and inactivating mutations on the phosphorylation and trafficking of the human lutropin/choriogonadotropin receptor. Mol Endocrinol 2000;14:1797–1810.
41. Latronico AC, Shinozaki H, Guerra G, Jr., Pereira MA, Lemos Marini SH, Baptista MT, Arnhold IJ, Fanelli F, Mendonca BB, Segaloff DL. Gonadotropin-independent precocious puberty due to luteinizing hormone receptor mutations in Brazilian boys: a novel constitutively activating mutation in the first transmembrane helix. J Clin Endocrinol Metab 2000;85:4799–4805.
42. Zeng FY, Hopp A, Soldner A, Wess J. Use of a disulfide cross-linking strategy to study muscarinic receptor structure and mechanisms of activation. J Biol Chem 1999;274:16629–16640.
43. Pauwels PJ, Wurch T, Tardif S, Finana F, Colpaert FC. Analysis of ligand activation of alpha 2-adrenoceptor subtypes under conditions of equal G alpha protein stoichiometry. Naunyn Schmiedebergs Arch Pharmacol 2001;363:526–536.
44. Kremer H, Martens JW, van Reen M, Verhoef-Post M, Wit JM, Otten BJ, Drop SL, Delemarre-van de Waal HA, Pombo-Arias M, De Luca F, Potau N, Buckler JM, Jansen M, Parks JS, Latif HA, Moll GW, Epping W, Saggese G, Mariman EC, Themmen AP, Brunner HG. A limited repertoire of mutations of the luteinizing hormone (LH) receptor gene in familial and sporadic patients with male LH-independent precocious puberty. J Clin Endocrinol Metab 1999;84:1136–1140.
45. Kristiansen K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. Pharmacol Ther 2004;103:21–80.
46. Mirzadegan T, Benko G, Filipek S, Palczewski K. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. Biochemistry 2003;42:2759–2767.
47. Ballesteros JA, Shi L, Javitch JA. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure–function analysis of rhodopsin-like receptors. Mol Pharmacol 2001;60:1–19.
48. 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.
49. Liapakis G, Ballesteros JA, Papachristou S, Chan WC, Chen X, Javitch JA. The forgotten serine: a critical role for Ser-203 5.42 in ligand binding to and activation of the beta 2-adrenergic receptor. J Biol Chem 2000;275:37779–37788.
50. Jiang Q, Guo D, Lee BX, Van Rhee AM, Kim YC, Nicholas RA, Schachter JB, Harden TK, Jacobson KA. A mutational analysis of residues essential for ligand recognition at the human P2Y1 receptor. Mol Pharmacol 1997;52:499–507.
51. Kokkola T, Foord SM, Watson MA, Vakkuri O, Laitinen JT. Important amino acids for the function of the human MT1 melatonin receptor. Biochem Pharmacol 2003;65:1463–1471.
52. Stitham J, Stojanovic A, Merenick BL, O'Hara KA, Hwa J. The unique ligand-binding pocket for the human prostacyclin receptor: site-directed mutagenesis and molecular modeling. J Biol Chem 2003;278:4250–4257.
53. Zhu S, Huys I, Dyason K, Verdonck F, Tytgat J. Evolutionary trace analysis of scorpion toxins specific for K-channels. Proteins 2004;54:361–370.
54. Shackelford GS, Regni CA, Beamer LJ. Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily. Protein Sci 2004;13:2130–2138.
55. Blaise MC, Sowdhamini R, Rao MR, Pradhan N. Evolutionary trace analysis of ionotropic glutamate receptor sequences and modeling the interactions of agonists with different NMDA receptor subunits. J Mol Model 2004;10:305–316
56. Innis CA, Shi J, Blundell TL. Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. Protein Eng 2000;13:839–847.
57. Xie T, Chen J, Ding DF. An evolutionary trace method for functional prediction of genomes. Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai) 1999;31:433–439.
58. Pritchard L, Dufton MJ. Evolutionary trace analysis of the Kunitz/BPTI family of proteins: functional divergence may have been based on conformational adjustment. J Mol Biol 1999;285: 1589–1607.
59. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 1996;257:342–358.
60. Townsend-Nicholson A, Schofield PR. A threonine residue in the seventh transmembrane domain of the human A_1 adenosine receptor mediates specific agonist binding. J Biol Chem 1994;269: 2373–2376.
61. Tucker AL, Robeva AS, Taylor HE, Holeton D, Bockner M, Lynch KR, Linden J. A_1 adenosine receptors: two amino acids are responsible for species differences in ligand recognition. J Biol Chem 1994;269:27900–27906.
62. Kim J, Wess J, van Rhee AM, Schoneberg T, Jacobson KA. Site-directed mutagenesis identifies residues involved in ligand recognition in the human A_{2A} adenosine receptor. J Biol Chem 1995;270:13987–13997.
63. Jiang Q, Van Rhee AM, Kim J, Yehle S, Wess J, Jacobson KA. Hydrophilic side chains in the third and seventh transmembrane helical domains of human A_{2A} adenosine receptors are required for ligand recognition. Mol Pharmacol 1996;50:512–521.
64. Dalpiaz A, Townsend-Nicholson A, Beukers MW, Schofield PR, IJzerman AP. Thermodynamics of full agonist, partial agonist, and antagonist binding to wild-type and mutant adenosine A_1 receptors. Biochem Pharmacol 1998;56:1437–1445.