PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

GGIP: Structure and sequence-based GPCR–GPCR interaction pair predictor

Wataru Nemoto,[1,2]* Yoshihiro Yamanishi,[3,4] Vachiranee Limviphuvadh,[5] Akira Saito,[1] and Hiroyuki Toh[2,6]

[1] Division of Life Science and Engineering, School of Science and Engineering, Tokyo Denki University (TDU), Ishizaka, Hatoyama-Machi, Hiki-Gun, Saitama, 350-0394, Japan

[2] Computational Biology Research Center (CBRC), Advanced Industrial Science and Technology (AIST), AIST Tokyo Waterfront Bio-IT Research Building, 2-4-7 Aomi, Koto-Ku, Tokyo, 135-0064, Japan

[3] Medical Institute of Bioregulation (MiB), Kyushu University, 3-1-1 Maidashi, Higashi-Ku, Fukuoka, 812-8582, Japan

[4] Institute for Advanced Study, Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan

[5] Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, #07-01 Matrix, 138671, Singapore

[6] Department of Biomedical Chemistry, School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda-Shi, Hyogo, 669-1337, Japan

ABSTRACT

G Protein-Coupled Receptors (GPCRs) are important pharmaceutical targets. More than 30% of currently marketed pharmaceutical medicines target GPCRs. Numerous studies have reported that GPCRs function not only as monomers but also as homo- or hetero-dimers or higher-order molecular complexes. Many GPCRs exert a wide variety of molecular functions by forming specific combinations of GPCR subtypes. In addition, some GPCRs are reportedly associated with diseases. GPCR oligomerization is now recognized as an important event in various biological phenomena, and many researchers are investigating this subject. We have developed a support vector machine (SVM)-based method to predict interacting pairs for GPCR oligomerization, by integrating the structure and sequence information of GPCRs. The performance of our method was evaluated by the Receiver Operating Characteristic (ROC) curve. The corresponding area under the curve was 0.938. As far as we know, this is the only prediction method for interacting pairs among GPCRs. Our method could accelerate the analyses of these interactions, and contribute to the elucidation of the global structures of the GPCR networks in membranes.

Proteins 2016; 84:1224–1233.

Key words: membrane proteins; protein–protein interaction; interaction partner prediction; support vector machine; drug design; computational biology.

INTRODUCTION

G Protein-Coupled Receptors (GPCRs) constitute the largest family of plasma membrane proteins. It is estimated that there are over 1000 GPCRs encoded in the human genome. Approximately 600 of the GPCRs are represented by sensory receptors and are thus considered to be therapeutically irrelevant, whereas the remaining 400 are potential drug targets.[1] Numerous studies have reported that GPCRs function not only as monomers but also as oligomers; that is, dimers or higher-order molecular complexes.[2–6] Many studies have reported that the functions of GPCRs, such as the ligand binding specificity, the affinity, and the downstream signaling activity, are modified through homo- and hetero-dimer formation.[3–5,7–9] Although their functional importance remains enigmatic and in some cases even controversial, as summarized by Terrillon,[7] several stages of the GPCR life cycle in a cell could be affected by dimerization (see Ref. 7 for review). In addition, GPCR oligomerizations

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Tokyo Denki University Science Promotion Fund; Grant number: Q13L-03; Grant sponsor: JSPS KAKENHI; Grant number: 25870764; Grant sponsors: Program to Disseminate Tenure Tracking System, MEXT, Japan and Kyushu University Interdisciplinary Programs in Education and Projects in Research Development; Grant sponsor: A*STAR JCO; Grant number: JCOAG04_FG03_2009; Grant sponsor: JSPS KAKENHI; Grant number: 24500362.

Akira Saito's current address is Software Service, 2-6-1, Nishi-Miyahara, Yodogawa-Ku, Osaka, 532-0004, Japan

*Correspondence to: Wataru Nemoto. E-mail: watarunemoto@gmail.com or w.nemoto@mail.dendai.ac.jp

Received 2 January 2016; Revised 26 April 2016; Accepted 9 May 2016
Published online 18 May 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.25071

are associated with various diseases.[5] For instance, the hetero-oligomerization between cannabinoid receptor 1 (Cbr1) and angiotensin receptor 1 (Agtr1) is suggested to be involved in the enhancement of angiotensin II-mediated signaling and the coupling of Agtr1 with multiple trimeric G proteins in hepatic stellate cells, in ethanol-treated rats with upregulated Cbr1.[10] Therefore, in liver fibrosis caused by alcohol consumption, the Cbr1-Agtr1 heterodimer may be a novel target for anti-fibrotic compounds.[10] Thus, studies of GPCR oligomerization are necessary not only to understand disease mechanisms but also to reveal novel pathways for the treatment of various illnesses (see ref. 5 for review). Oligomer-specific ligands are now considered to be one of the promising strategies to control GPCR functions precisely.[8,9,11] Devi *et al.* identified compounds that selectively activate the μ-δ opioid receptor heterodimer, which thus have potential uses in pain treatment without the undesirable effects associated with opiate use (PubChem AID 588407).[12]

If we suppose that there are 1000 GPCR genes in the human genome, then the number of all possible hetero pairs from the 1000 GPCRs is 499,500. Systematic experimental studies of oligomerization for such a vast number of combinations of GPCRs would be time-consuming and costly. Hence, an accurate method to predict interacting pairs could accelerate such studies, by reducing their costs.

There are various methods to predict interacting pairs between soluble proteins[13] or between soluble and membrane proteins.[14] However, these methods cannot be applied to predict interacting GPCR pairs, since the characteristics used for the predictions between soluble proteins or between soluble and membrane proteins are different from those between membrane proteins.

We have developed GPCR–GPCR interacting pair predictor (GGIP), a method to predict specifically interacting pairs for GPCR hetero-oligomerization, by integrating the structure and sequence information of GPCRs. In GGIP, different structural regions are assumed to be used for the interaction interfaces among different GPCRs, because the interfaces for GPCR–GPCR interactions are not always conserved even among closely related GPCRs.[15–17] The prediction by GGIP is made by the support vector machine (SVM), which classifies a given hetero GPCR pair into either an interacting pair or a non-interacting pair by using a pair-wise kernel. In this article, we describe the advantages and disadvantages of our method.

MATERIALS AND METHODS

Preparation of Training Data and Prediction Targets

In this study, we used 353 mouse non-odorant GPCR sequences with available gene expression profiles in 41 tissues,[18] as the prediction targets. These target amino acid sequences were retrieved from the RefSeq database.[19,20] The RefSeq IDs of the mouse sequences are provided in Supporting Information Table SI. We used the GPCR abbreviations reported in Regard's work.[18] Their homologous sequences in human are the principal targets for medical treatments. The number of all possible pairs of the 353 GPCRs is 62,128 ($_{353}C_2$).

Among them, 61 pairs were treated as positive data for training, to develop the predictor. Among the 61 pairs, 16 pairs are registered as mouse non-odorant GPCR pairs in the GPCR oligomer databases.[21,22] Referring to the databases and the literature,[21–34] the remaining 45 pairs were identified in non-mouse species, but their orthologous pairs are present in mice. All positive pairs are listed in Supporting Information Table SII.

Negative data candidates for training to develop the predictor were collected, as follows. As negative data candidates, 21 pairs of mouse receptors with orthologous pairs from other species that have been experimentally characterized as non-interactive with each other[4,21,23,29,35–39] were used. To compensate for the imbalance in the sizes of the positive and negative datasets, the negative dataset was augmented as follows. Among the 62,128 possible GPCR pairs, 9275 pairs are not coexpressed in any of the 41 examined mouse tissues.[18] The 9275 pairs were also used as the negative data candidates, under the assumption that a pair of proteins, which are not coexpressed, will not interact with each other. By adding the 9275 pairs to the 21 negative pairs, the total number of negative data candidates became 9296 pairs (= 21 + 9275). The negative data candidates are listed in Supporting Information Table SIII.

The negative pairs for the training were randomly selected from the negative data candidates. The size of the negative data set thus selected was set to three times greater than the number of positive pairs. Then, the experimentally suggested 21 pairs were always selected as the negative data.

The interactions of the remaining 52,771 pairs (= 62,128 – 61 – 9296) have not been characterized yet, although they are coexpressed in at least one of the 41 tissues. These pairs were used to predict novel interacting pairs.

SVM-Based Method to Predict GPCR–GPCR Interacting Pairs

To predict GPCR–GPCR interaction pairs, we used the SVM. In recent years, the SVM has been gaining popularity in the analyses of biological problems,[40–45] such as gene and tissue classifications from microarray expression data,[41] protein subcellular localization prediction,[40,42] protein function classification,[44] and protein structural class prediction.[45] By learning a set of positively and negatively labeled training samples, the SVM classifies new unlabeled test samples.

In this study, we regarded the pair of the x and y GPCRs as the sample, and used the SVM to classify GPCR–GPCR pairs into either the interaction or non-interaction class. Suppose that we have a set of GPCR–GPCR pairs $\{(x, y)_1, (x, y)_2, \ldots, (x, y)_m\}$, where m is the number of GPCR–GPCR pairs in the learning set, and each pair in the set belongs to either the interaction or non-interaction class. The SVM classifier is formulated as follows:

$$f(x', y') = \sum_{i=1}^{m} \tau_i K_{pair}\left((x, y)_i, (x', y')\right) \quad (1)$$

where (x', y') is any new GPCR–GPCR pair to be classified, K_{pair} is a kernel similarity function for GPCR–GPCR pairs, and $\{\tau_1, \tau_2, \ldots, \tau_m\}$ are the parameters learned. If $f(x', y')$ is positive, then GPCR x' and GPCR y' are predicted to interact with each other. In contrast, if $f(x', y')$ is negative, then GPCR x' and GPCR y' are predicted not to interact. We compute the pairwise kernel function for the GPCR–GPCR pairs in the SVM algorithm, as follows:

$$K_{pair}((x, y), (x', y')) = K(x, x')K(y, y') + K(x, y')K(x', y) \quad (2)$$

where K is a kernel similarity function for GPCRs. The pairwise kernel function was proposed in a previous study for predicting protein–protein interactions.[45]

We prepared a set of values (e.g., 0.001,0.01, 0.1,1,10,100,1000) for the regularization parameter C in SVM, tested each value to calculate the classification accuracy by cross-validation, and chose the value that provided the highest accuracy. The best parameter value depends on a random seed used for data splitting in the cross-validation, but the accuracy tends to be the highest when the regularization parameter C is 1. Thus, we set $C = 1$ in the analyses in the article.

Preparation of Feature Vectors

The feature vectors for the pairs were generated by integrating the structure and sequence information, as follows.

Preprocessing of Structure and Sequence Data of Each GPCR

To obtain the structure information for each target GPCR sequence, the structure of the most closely related sequence to the target was used as the template. At present, 125 PDB entries contain GPCR crystal structures belonging to the class A, B, or C GPCRs. When more than two Protein Data Bank (PDB) entries contain an identical GPCR, we adopted the entry of the protein structure with the highest resolution and the maximum sequence length as the candidate for the template structure.

We used the chain where transmembrane regions are annotated in the Protein Data Bank of Transmembrane Proteins (PDBTM).[46] As a result, 65 structures were adopted as candidates for a template structure for each GPCR. Among them, the structure of the most closely related sequence to the target GPCR sequence was selected. The adopted PDB IDs and their chains are listed in Supporting Information Table SIV.

In each PDB file, we removed the hetero atom (HETATM) lines. The ATOM lines corresponding to an artificially conjugated domain that was used to facilitate the crystallization of a PDB structure were removed. For instance, the atom (ATOM) lines corresponding to lysozyme, which is artificially connected to intracellular loop III of β_2 adrenergic receptor (Adrb2) (PDB ID: 3D4S chain A),[47] were removed.

Segmentation of a GPCR Sequence Based on the Template Structure

There are two possible mechanisms for the formation of a GPCR oligomer. One is domain contact oligomerization, while the other is domain-swapping oligomerization.[48] Although the two mechanisms are still under debate, there is some evidence to support the domain contact mechanism.[49–53] Therefore, we have developed a method based on the assumption that GPCRs form oligomers by the domain contact mechanism. Domain contact oligomers reportedly have at least one of three types of interaction interfaces for oligomerization.[54] The first type of interface exists in the amino-terminal extracellular region, which forms disulfide bonds with the corresponding region of an interaction partner.[55] The second interface type is at the carboxyl cytoplasmic tail, which forms a coiled-coil structure with an interaction partner.[56] The third type is located on the lipid-facing surface along the seven transmembrane helices. The critical residues for oligomerization seem to be scattered along the transmembrane helices[15,51,52,57] or loops,[58] but have the loose regulation to cluster on either the extracellular side, central or intracellular side of transmembrane helices, or at some of them.[15,51,52] The details of the oligomerizations based on these mechanisms are provided in the Supporting Information.

Based on the observations of the domain contact oligomerizations, we divided each target GPCR sequence into 29 segments, with reference to the structural information of the template structure (Fig. 1). First, the N- and C-termini, the three extracellular loops, and the three intracellular loops constitute eight segments. Second, each transmembrane helix was further split into three segments: the seven residues constituting the N-terminus of the helical region, the seven residues constituting the C-terminus of the helical region, and the remaining region constituting the center of the helix.

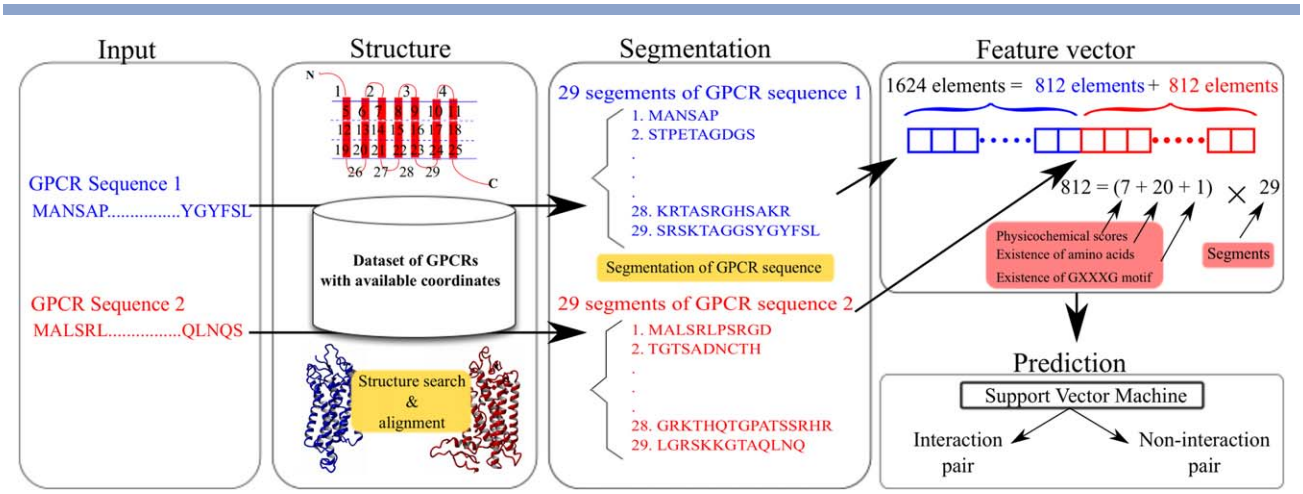

Figure 1

Procedure for the development of our method. Two amino acid sequences of a hetero GPCR pair were used as input. Each sequence was split into 29 segments, according to a closely related structure. The average scores of various characteristics were calculated for each segment, and used as the elements of a feature vector for a GPCR pair.

Then, the helical region that is continuous with a transmembrane helix, but located outside the membrane, is included in one of the loops. Thus, the transmembrane helices were split into 21 ($= 3 \times 7$) segments.

As a result, 29 ($= 8 + 21$) segments were obtained. A label corresponding to one of the structural segments was assigned to each residue of a target GPCR sequence, based on the pairwise alignment between the sequence of the target GPCR and that of the template structure. We focused on the exposed residues in each segment, because we assumed in this study that buried residues are not involved in the physical contact for the GPCR–GPCR interaction. An exposed residue was defined as a residue with a relative accessible surface area (rASA) $\geq$25%.[15] The rASA was calculated by using NACCESS.[59]

Designing of a Feature Vector for Each Segment

As shown in Figure 1, the feature vector for each pair consists of 1624 elements, where two sets of 812 scores corresponding to target GPCRs are integrated ($1624 = 2 \times 812$). The 812 scores consist of 28 scores for the defined 29 segments of each GPCR ($812 = 28 \times 29$).

To represent the features of each segment, 28 scores were calculated. Among them, seven scores are the averages of three hydrophobicity scores,[60–62] the log of partition coefficient ($\log P$),[63] the acidity,[64] the polarity,[64] and the isoelectric point (pI)[64] over the amino acid residues in each segment. The remaining 21 scores were calculated to indicate the presence or absence of 20 types of amino acids and the GXXXG motif[65] in each segment. These 21 kinds of presence or absence scores were represented as binary expressions.

As a result, a single GPCR had 812 scores ($= (7 + 21) \times 29$). The feature vector for each pair is designed to integrate the two sets of 812 scores corresponding to the pair of target GPCRs. The obtained vector has the size of 1624 elements.

Performance Evaluations

In this study, the performance of a predictor was evaluated with a Receiver Operating Characteristic (ROC) curve. The horizontal and vertical axes of the ROC curve represent the false positive and true positive rates, respectively. Both rates were obtained by varying the threshold used to classify interacting and non-interacting GPCR pairs. Herein, a true positive is a correctly predicted pair, whereas a false positive is a predicted pair that is not present in the positive dataset.

The prediction performance was evaluated based on the Area Under the ROC Curve (AUC).[66] The value of AUC ranges from 0 to 1. An AUC value close to 1 indicates the good performance of the predictor. The 5-fold cross validation was performed for a predictor, as follows. The set of all pairs was split into five equally sized subsets. In each validation cycle, one subset was removed for evaluation, and the remaining subsets were used for the training of the predictor. Thus, five AUCs were obtained for each cross-validation cycle. The average of the five AUCs was used as a measure of the performance of a predictor. Performance comparisons of the five predictors are provided in Figure 2.

Other Predictors With Different Feature Vectors

We evaluated the prediction accuracy of GGIP, by comparisons with those of the four other predictors. Details of the other predictors are described below.

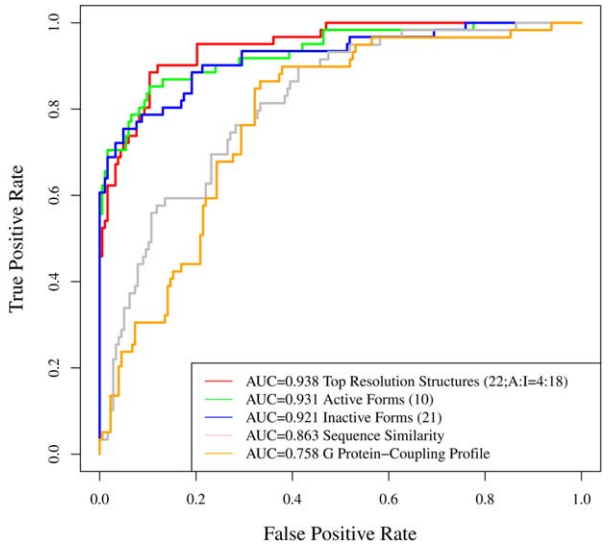

Figure 2

Performance comparison of GGIP and other predictors.

Other Structure- and Sequence-Based Predictors

Some GPCRs reportedly form oligomers in an active (agonist-bound) conformation, while others are in an inactive (antagonist-bound) conformation.[3,67,68] To examine whether the conformation of a template structure affects the prediction accuracy of GGIP, we prepared two other structure- and sequence-based predictors that use either an active or inactive conformation structure as the template.[69] The active-form predictor uses one of the 10 agonist-bound structures as the template of each target GPCR, while the inactive form predictor uses one of the 55 antagonist-bound structures (Supporting Information Table SV). The procedures used to generate their feature vectors are the same as those of GGIP.

Sequence-Based Predictor

Ben-Hur *et al.*[45] proposed the use of sequence similarity for predicting protein–protein interactions in the SVM framework.[70] Here, we also tested the sequence similarity for the prediction of interacting GPCR pairs. The sequence similarity score S_{ij} is calculated as a kernel similarity function of GPCRs, as follows:

$$S_{ij} = Smith\text{-}Waterman\ \ Score/\left(N_i \times N_j\right) \qquad (3)$$

where the Smith-Waterman based score[70] is divided by the lengths of sequences i and j (N_i and N_j). The resulting normalized score is transformed into a positive definite kernel by subtracting the minimum eigenvalue. Then, the corresponding pairwise kernel is computed for all of the pairs among the 353 GPCRs.

G Protein-Coupling Profile-Based Predictor

Most GPCRs exert their signal transduction activity through coupling with one or several types of G proteins, although a few GPCRs have G protein-independent functions.[71] The G protein types are specific for a particular set of GPCRs and downstream signaling pathways. We tested the profile of the coupled G proteins for the prediction of interacting GPCR pairs. The set of the coupled G proteins of a GPCR was expressed as a four-dimensional binary vector. Each element of the vector corresponds to G_s, G_i, G_q, and $G_{o/11}$. If the GPCRs coupled with a G protein, then the corresponding element of the vectors is set to 1. Otherwise, it is 0. Information about the coupled G proteins of 245 target GPCRs was manually retrieved from the International Union of Basic and Clinical Pharmacology (IUPHAR) website.[72] The coupled G proteins for the remaining 108 GPCRs were predicted by Griffin,[14] which is a web-based system to predict a coupled G protein from the amino acid sequence of a given GPCR. The four-dimensional profile was used as a feature vector of each GPCR, and the similarity of the GPCR profiles was calculated by the Gaussian RBF kernel. Then, the corresponding pairwise kernel based on the profile similarities was computed for all of the pairs among the 353 GPCRs.

RESULTS

The performance of GGIP was evaluated by an ROC curve (Fig. 2). The AUC[66] was 0.938 (red line). We also tested other features for the prediction accuracy. The AUCs of these predictors were 0.931 (green: active conformation and sequence information), 0.921 (blue: inactive conformation and sequence information), 0.863 (orange: sequence similarity), and 0.758 (grey: G protein-coupling profile), respectively. Comparisons with the four approaches revealed that GGIP showed the highest performance, although the true positive rates of the predictors with an active or inactive conformation are larger than that of GGIP in the region where the false positive rate is smaller than 0.100. All of the predicted interacting pairs are listed in Supporting Information Table SVI. In addition to 5-fold cross-validation, we performed leave-one-out cross-validation, and confirmed that the proposed method tends to work the best, regardless of the cross-validation procedure. We evaluated the effects of data imbalance by varying the ratio between positive and negative samples (e.g., 1:1, 1:2, 1:3, 1:4, 1:5). The corresponding AUC scores are summarized in the Supporting Information Table SVII. These results suggest that the proposed method tends to work the best, regardless of data imbalance. In an attempt to improve the prediction accuracy of GGIP, these types of information were integrated in various combinations and

used to design predictors. However, the performance was not improved.

DISCUSSION

Advantages and Disadvantages of the Predictors

As described above, GGIP (AUC = 0.938) outperformed the other predictors (Fig. 2). The active conformation-based predictor (AUC = 0.931) and the inactive conformation-based predictor (AUC = 0.921) also showed high performance. In contrast, the performances of the predictors without structure information were comparatively low. These results imply that structure information contributes to the high performance of the predictor. As described in the Materials and methods section, three structure-based predictors differ in the conformations of the template structure candidates. GGIP adopted the structure of the most closely related sequence as the template of the target GPCR, regardless of the conformation. As a result, the adopted structures for GGIP consist of four structures in an active conformation and 18 structures in an inactive conformation. The active conformation-based predictor and the inactive conformation-based predictor used 10 structures in an active conformation and 21 structures in an inactive conformation, respectively. Comparisons of the performances of the three structure-based predictors showed that the conformational differences among the template structures did not cause significant differences in the performance. This may signify that the difference in the amino acid composition of a structural segment between closely related sequences with different conformations is smaller than that between distantly related sequences with the same conformations. The result suggests that the accuracy of the prediction is improved if the structures with amino acid sequences that are close to those of the GPCR pair under consideration are available, regardless of their conformations. Therefore, an increase in the available coordinates of GPCRs would further improve the accuracy of our prediction method.

The sequence similarity-based predictor demonstrated an AUC of 0.863. Although this performance is lower than those of the structure-based predictors, the good performance of this predictor seems to agree with the tendency that proteins with high sequence similarity interact with each other.[23] However, GPCRs with high sequence similarity do not always interact with each other.[23] For example, μ-opioid receptor (Oprm) interacts with δ-opioid receptor (Oprd) and κ-opioid receptor (Oprk), but Oprd does not form an oligomer with Oprk.[11,38,73,74] The percent sequence identities between the human opioid receptor subtypes are 59% (Oprk–Oprd), 60% (Oprm–Oprk), and 63% (Oprm–Oprd). Thus, sequence similarity and identity do not seem to have a relationship with GPCR–GPCR interactions. This might be the reason for the lower performance of the sequence-based predictor.

The AUC for the G protein coupling profile-based predictor was 0.758, which was the lowest among all of the predictors. This result indicates that two GPCRs with identical G protein coupling profiles do not always interact with each other. In fact, the oligomerization can occur between GPCRs that couple with different G proteins. For example, G_i/G_o-coupled A1 adenosine receptor and G_q-coupled P2Y1 receptor coimmunoprecipitate in cotransfected HEK293T cells.[75] In addition, several GPCRs reportedly change their coupling G protein types through the formation of heterodimers. For example, G_q/G_{11}-coupling growth hormone secretagogue receptor 1b and neurotensin receptor 1 heterodimerize in non-small cell lung cancer cells. This heterodimer binds neuromedin U and couples with G_s, although the endogenous neuromedin U receptor couples with G_q/G_{11}.[76] Thus, the G protein coupling profile seems to be more complicated than previously thought. Further consideration will be needed to use the profile for the development of a high performance predictor.

Disease-Related Interacting GPCR Pairs

As described above, various diseases are associated with specific combinations of GPCRs.[5] A recent study reported that somatic mutations in GPCR-encoding genes are frequently found in many types of cancer.[77] Among the somatic mutations, hotspot mutations are defined as recurrent amino acid changes occurring in coding sequences. The hotspot mutations observed in G proteins and GPCRs were recently summarized in a review.[78] The data revealed that many GPCRs have hotspot mutations in the same types of cancer tissues (see Supporting Information Table SIII of this review[78]). For example, 145 GPCRs with hotspot mutations were found in lung cancer tissues.[78]

Most of the hotspot mutations have not been characterized yet, and their effects on the cancer pathways remain unknown.[78] We expected to find some examples of mutations that may cause cancers by modifying GPCR oligomerization. Hence, we examined the 415 predicted interacting GPCR pairs, to determine whether there are GPCRs with hotspot mutations in the same types of cancer tissues. Among them, 73 predicted interacting pairs present in the same types of cancer tissues have hotspot mutations in each monomer of the predicted pairs. We then examined whether these hotspot mutations are present in the predicted interfaces of the GPCRs by using the method to predict interfaces for GPCR oligomerization,[16,79] which predicts at most two interface regions for a given GPCR. The interface prediction requires the coordinate data of a GPCR monomer structure,[15,79] and this restriction further reduced the number of the

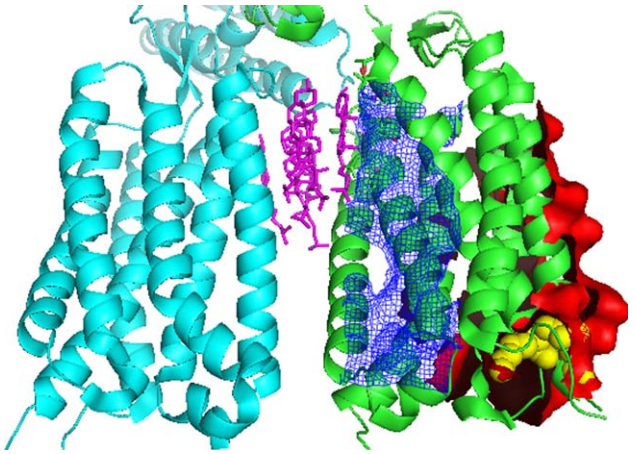

Figure 3

Two predicted interfaces mapped on a protomer constituting a homodimer of Grm1. Two predicted interfaces are indicated in red smoothed surface and blue mesh, respectively. The residues corresponding to the hotspot mutations R681H/G and R684C are depicted as yellow spheres.

GPCRs we could examine. We found hotspot mutations in Glutamate receptor type 1 (Grm1), as a possible example that may be related to the pathology of lung cancer by affecting the heterodimerization with Adrenergic receptor A2b (Adra2b). Both GPCRs have hotspot mutations associated with lung cancer. Although no coordinate data of Adra2b are currently available, the structure of the Grm1 homodimer was recently determined.[80] Therefore, we predicted the interfaces of Grm1 in this study.

Two predicted interfaces were mapped on chain A of the crystal structure of the Grm1 homodimer[80] (see Fig. 3). One of the interfaces consisted of helix II and the extracellular side of helix III, while the other interface was formed by the membrane-facing surface regions of helices IV and V, along with the cytoplasmic tail of helix III. As shown in Figure 3, the former interface agreed with the interface for homodimerization.[80] Therefore, the latter interface could be used for the interaction with Adra2b. Among the hotspot mutations in the Grm1, R681H/G and R684C were found on the cytoplasmic surface of helix III, one of the constituents of the interface for heterooligomerization (depicted as yellow spheres in Fig. 3). These hotspot mutations might be related to the pathology of lung cancer, by affecting heterodimerization.

As described above, many aspects of the interactions among GPCRs remain to be elucidated. High performance prediction methods could accelerate the analyses of these interactions, and contribute to the elucidation of the global structures of the GPCR networks in membranes. Thus, the integration of mutation data and the predicted interacting pairs could shed light on how hotspot mutations may influence the GPCR networks and affect the mechanisms related to the pathology of cancer.

Concluding Remarks

We have developed a high-performance method to predict interacting GPCR pairs. We would like to conclude by describing possible improvements and extensions of our method.

The number of target GPCRs in this work was restricted to the targets in Regard's work,[18] because the current version of GGIP requires comprehensive gene expression data of the targets to generate negative data candidates. However, the number of targets will increase when we obtain more gene expression data or integrate the gene expression data of other GPCRs available from public databases.

Membrane lipids have recently been implicated in the modulation of GPCR oligomerization.[81–84] In particular, the roles of membrane cholesterol content have been extensively investigated. Ganguly *et al.*[83] showed that cholesterol depletion appears to increase the dimeric fraction of the receptor population. Coarse-grained molecular dynamics simulations showed that the composition of the helices at the interface of the dimer for the Adrb2 varied, depending on the membrane cholesterol content.[85] Considering that cellular cholesterol is developmentally regulated and its content increases with aging,[86,87] Chattopadhyay advocated that the organization of GPCR oligomers could also be age-dependent.[81] In the next version of GGIP, we need to consider the cholesterol-mediated interactions between GPCR molecules, and to reconstruct positive and negative datasets with age-dependent GPCR oligomerization.

The current version of GGIP cannot perform prediction when the structure information of query GPCRs is unavailable. However, secondary structure prediction and molecular surface prediction could compensate for the lack of structure information. The next version of GGIP will perform prediction without structure information of query GPCRs.

ACKNOWLEDGMENTS

We thank S. Maurer-Stroh for valuable comments about the relationships between the interacting GPCR pairs and the disease-related nsSNPs.

AUTHOR CONTRIBUTIONS

W.N. and H.T. conceived the concept to develop the prediction method and designed the feature vectors. Y.Y. optimized the parameters of the SVM procedures. V.L. and W.N. investigated the relationships between the interacting GPCR pairs and the disease-related nsSNPs.

A.S. extended the method to predict interfaces for GPCR oligomerization for this research. All authors contributed to write the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Mcdonald P. In-depth focus new opportunities and challenges for drug discovery. Eur Pharm Rev 2012;17:2–13.
2. Rivero-Müller A, Chou Y-Y, Ji I, Lajic S, Hanyaloglu AC, Jonas K, Rahman N, Ji TH, Huhtaniemi I. Rescue of defective G protein-coupled receptor function in vivo by intermolecular cooperation. Proc Natl Acad Sci 2010;107:2319–2324.
3. Milligan G. The role of dimerization in the cellular trafficking of G-protein-coupled receptors. Curr Opin Pharmacol 2010;10:23–29.
4. Nemoto W, Toh H. Membrane interactive alpha-helices in GPCRs as a novel drug target. Curr Protein Pept Sci 2006;7:561–575.
5. Gomes I, Fujita W, Chandrakala MV, Devi LA. Disease-specific heteromerization of G-protein-coupled receptors that target drugs of abuse. Prog Mol Biol Transl Sci 2013;117:207–265.
6. Kasai RS, Kusumi A. Single-molecule imaging revealed dynamic GPCR dimerization. Curr Opin Cell Biol 2014;27:78–86.
7. Terrillon S, Bouvier M. Roles of G-protein-coupled receptor dimerization. EMBO Rep 2004;5:30–34.
8. Bellot M, Galandrin S, Boularan C, Matthies HJ, Despas F, Denis C, Javitch J, Mazéres S, Sanni SJ, Pons V, Seguelas MH, Hansen JL, Pathak A, Galli A, Sénard JM, Galés C. Dual agonist occupancy of AT1-R–α2C-AR heterodimers results in atypical Gs-PKA signaling. Nat Chem Biol 2015;11:271–279.
9. Parmentier M. GPCRs: Heterodimer-specific signaling. Nat Chem Biol 2015;11:244–245.
10. Rozenfeld R, Gupta A, Gagnidze K, Lim MP, Gomes I, Lee-Ramos D, Nieto N, Devi LA. AT1R-CB_1R heteromerization reveals a new mechanism for the pathogenic properties of angiotensin II. Embo J 2011;30:2350–2363.
11. Gomes I, Jordan BA, Gupta A, Trapaidze N, Nagy V, Devi LA. Heterodimerization of mu and delta opioid receptors: a role in opiate synergy. J Neurosci 2000;20:RC110.
12. Lakshmi A. Devi. Luminescence-based cell-based high throughput dose response assay for agonists of heterodimerization of the mu 1 (OPRM1) and delta 1 (OPRD1) opioid receptors (PubChem AID 588407), 2011.
13. Pitre S, Alamgir M, Green JR, Dumontier M, Dehne F, Golshani A. Computational methods for predicting protein–protein interactions. Adv Biochem Eng Biotechnol 2008;110:247–267.
14. Yabuki Y, Muramatsu T, Hirokawa T, Mukai H, Suwa M. GRIFFIN: a system for predicting GPCR-G-protein coupling selectivity using a support vector machine and a hidden Markov model. Nucleic Acids Res 2005;33:W148–W153.
15. Nemoto W, Toh H. Prediction of interfaces for oligomerizations of G-protein coupled receptors. Proteins 2005;58:644–660.
16. George SR, Lee SP, Varghese G, Zeman PR, Seeman P, Ng GYOB. A transmembrane domain-derived peptide inhibits D1dopamine receptor function without affecting receptor oligomerization. J Biol Chem 1998;273:30244–30248.
17. Soyer OS, Dimmic MW, Neubig RRGR. Dimerization in aminergic G-protein-coupled receptors: application of a hidden- site class model of evolution. Biochemistry 2003;42:14522–14531.
18. Regard JB, Sato IT, Coughlin SR. Anatomical profiling of G protein-coupled receptor expression. Cell 2008;135:561–571.
19. Pruitt K, Brown G, Tatusova T, and Maglott D. The NCBI handbook [Internet], Chapter 18, The Reference Sequence (RefSeq) Project. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; Oct 2002.
20. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI reference sequences (RefSeq): current status, new features, and genome annotation policy. Nucleic Acids Res 2012;40:D130–D135. Database issue):–.
21. Nemoto W, Fukui K, Toh H. GRIPDB – G protein coupled receptor interaction partners database. J Recept Signal Transduct Res 2011; 31:199–205.
22. Khelashvili G, Dorff K, Shan J, Camacho-Artacho M, Skrabanek L, Vroling B, Bouvier M, Devi LA, George SR, Javitch JA, Lohse MJ, Milligan G, Neubig RR, Palczewski K, Parmentier M, Pin JP, Vriend G, Campagne F, Filizola M. GPCR-OKB: the G protein coupled receptor oligomer knowledge base. Bioinformatics 2010;26:1804–1805.
23. Doumazane E, Scholler P, Zwier JM, Trinquet E, Rondard P, Pin J-P. A new approach to analyze cell surface protein complexes reveals specific heterodimeric metabotropic glutamate receptors. Faseb J 2011; 25:66–77.
24. Gama L, Wilt SG, Breitwieser GE. Heterodimerization of calcium sensing receptors with metabotropic glutamate receptors in neurons. J Biol Chem 2001;276:39053–39059.
25. Ramsay D, Kellett E, McVey M, Rees S, Milligan G. Homo- and hetero-oligomeric interactions between G-protein-coupled receptors in living cells monitored by two variants of bioluminescence resonance energy transfer (BRET): hetero-oligomers between receptor subtypes form more efficiently than between less. Biochem J 2002; 365:429–440.
26. Chen C, Li J, Bot G, Szabo I, Rogers TJ, Liu-Chen L-Y. Heterodimerization and cross-desensitization between the mu-opioid receptor and the chemokine CCR5 receptor. Eur J Pharmacol 2004;483:175–186.
27. Nickolls SA, Maki RA. Dimerization of the melanocortin 4 receptor: a study using bioluminescence resonance energy transfer. Peptides 2006;27:380–387.
28. Rios C, Gomes I, Devi LA. Mu opioid and CB1 cannabinoid receptor interactions: reciprocal inhibition of receptor signaling and neuritogenesis. Br J Pharmacol 2006;148:387–395.
29. Zaslavsky A, Singh LS, Tan H, Ding H, Liang Z, Xu Y. Homo- and hetero-dimerization of LPA/S1P receptors, OGR1 and GPR4. Biochim Biophys Acta 2006;1761:1200–1212.
30. Gehlert DR, Schober DA, Morin M, Berglund MM. Co-expression of neuropeptide Y Y1 and Y5 receptors results in heterodimerization and altered functional properties. Biochem Pharmacol 2007;74: 1652–1664.
31. Schröder H, Wu DF, Seifert A, Rankovic M, Schulz S, Höllt V, Koch T. Allosteric modulation of metabotropic glutamate receptor 5 affects phosphorylation, internalization, and desensitization of the micro-opioid receptor. Neuropharmacology 2009;56:768–778.
32. Kuhn C, Bufe B, Batram C, Meyerhof W. Oligomerization of TAS2R bitter taste receptors. Chem Senses 2010;35:395–406.
33. Li Y, Chen J, Bai B, Du H, Liu Y, Liu H. Heterodimerization of human apelin and kappa opioid receptors: roles in signal transduction. Cell Signal 2012;24:991–1001.
34. Delille HK, Becker JM, Burkhardt S, Bleher B, Terstappen GC, Schmidt M, Meyer AH, Unger L, Marek GJ, Mezler M. Heterocomplex formation of 5-HT2A-mGlu2 and its relevance for cellular signaling cascades. Neuropharmacology 2012;62:2184–2191.
35. Terrillon S, Durroux T, Mouillac B, Breit A, Ayoub MA, Taulan M, Jockers R, Barberis C, Bouvier M. Oxytocin and vasopressin V1a and V2 receptors form constitutive homo- and heterodimers during biosynthesis. Mol Endocrinol 2003;17:677–691.
36. Wang D, Sun X, Bohn LM, Sadée W. Opioid receptor homo- and heterodimerization in living cells by quantitative bioluminescence resonance energy transfer. Mol Pharmacol 2005;67: 2173–2184.

37. Rocheville M, Lange DC, Kumar U, Sasi R, Patel RC, Patel YC. Subtypes of the somatostatin receptor assemble as functional homo- and heterodimers. J Biol Chem 2000;275:7862–7869.
38. Jordan BA, Cvejic S, Devi LA. Opioids and their complicated receptor complexes. Neuropsychopharmacology 2000;23:S5–S18.
39. González S, Moreno-Delgado D, Moreno E, Pérez-Capote K, Franco R, Mallol J, Cortés A, Casadó V, Lluís C, Ortiz J, Ferré S, Canela E, McCormick PJ. Circadian-related heteromerization of adrenergic and dopamine D_4 receptors modulates melatonin synthesis and release in the pineal gland. PLoS Biol 2012;10:e1001347.
40. Chou KC, Elrod DW. Protein subcellular location prediction. Protein Eng 1999;12:107–118.
41. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 2000;16:906–914.
42. Park K-J, Kanehisa M. Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. Bioinformatics 2003;19:1656–1663.
43. Cai CZ, Wang WL, Sun LZ, Chen YZ. Protein function classification via support vector machine approach. Math Biosci 2003;185:111–122.
44. Cai Y-D, Liu X-J, Xu X, Chou K-C. Prediction of protein structural classes by support vector machines. Comput Chem 2002;26:293–296.
45. Ben-Hur A, Noble WS. Kernel methods for predicting protein–protein interactions. Bioinformatics 2005;21 Suppl 1:i38–i46.
46. Kozma D, Simon I, Tusnády GE. PDBTM: protein data bank of transmembrane proteins after 8 years. Nucleic Acids Res 2013;41: D524–D529.
47. Hanson MA, Cherezov V, Griffith MT, Roth CB, Jaakola VP, Chien EY, Velasquez J, Kuhn P, Stevens RC. A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. Structure 2008;16:897–905.
48. Gouldson PR, Snell CR, Bywater RP, Higgs C, Reynolds CA. Domain swapping in G-protein coupled receptor dimers. Protein Eng 1998;11:1181–1193.
49. Huang J, Chen S, Zhang JJ, Huang X-Y. Crystal structure of oligomeric β1-adrenergic G protein-coupled receptors in ligand-free basal state. Nat Struct Mol Biol 2013;20:419–425.
50. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 2010;330: 1066–1071.
51. Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K. Atomic-force microscopy: Rhodopsin dimers in native disc membranes. Nature 2003;421:127–128.
52. Hebert TE, Moffett S, Morello JP, Loisel TP, Bichet DG, Barret C, Bouvier M. A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. J Biol Chem 1996;271:16384–16392.
53. Zeng F, Wess J. Molecular aspects of muscarinic receptor dimerization. Neuropsychopharmacology 2000;23:S19–S31.
54. Bouvier M. Oligomerization of G-protein-coupled transmitter receptors. Nat Rev Neurosci 2001;2:274–286.
55. Tsuji Y, Shimada Y, Takeshita T, Kajimura N, Nomura S, Sekiyama N, Otomo J, Usukura J, Nakanishi S, Jingami H. Cryptic dimer interface and domain organization of the extracellular region of metabotropic glutamate receptor subtype 1. J Biol Chem 2000;275:28144–28151.
56. Kaupmann K, Malitschek B, Schuler V, Heid J, Froestl W, Beck P, Mosbacher J, Bischoff S, Kulik A, Shigemoto R, Karschin A, Bettler B. GABA(B)-receptor subtypes assemble into functional heteromeric complexes. Nature 1998;396:683–687.
57. Romano C, Miller JK, Hyrc K, Dikranian S, Mennerick S, Takeu-chi Y, Goldberg MPOK. Covalent and noncovalent interactions mediate metabotropic glutamate receptor mGlu5 dimerization. Mol Pharmacol 2001;59:46–53.
58. O'Callaghan K, Kuliopulos A, Covic L. Turning receptors on and off with intracellular pepducins: new insights into G-protein-coupled receptor drug development. J Biol Chem 2012;287:12787–12796.
59. Hubbard SJ, Thornton JM. NACCESS. 1993. Available at: http://www.bioinf.manchester.ac.uk/naccess/.
60. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
61. Eisenberg D. Three-dimensional structure of membrane and surface proteins. Annu Rev Biochem 1984;53:595–623.
62. Rose GD, Wolfenden R. Hydrogen bonding, hydrophobicity, packing, and protein folding. Annu Rev Biophys Biomol Struct 1993;22: 381–415.
63. V, Pliška M, Schmidt J-LF. Partition coefficients of amino acids and hydrophobic parameters π of their side-chains as measured by thin-layer chromatography. J Chromatogr A 1981;216:79–92.
64. Amino Acid Explorer. Available at: http://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi.
65. Senes A, Ubarretxena-Belandia I, Engelman DM. The Calpha —H. . .O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. Proc Natl Acad Sci 2001;98:9056–9061.
66. Gribskov M, Robinson NL. Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. Comput Chem 1996;20:25–33.
67. Fuxe K, Agnati LF, Jacobsen K, Hillion J, Canals M, Torvinen M, Tinner-Staines B, Staines W, Rosin D, Terasmaa A, Popoli P, Leo G, Vergoni V, Lluis C, Ciruela F, Franco R, Ferré S. Receptor heteromerization in adenosine A2A receptor signaling: relevance for striatal function and Parkinson's disease. Neurology 2003;61:S19–S23.
68. Grant M, Collier B, Kumar U. Agonist-dependent dissociation of human somatostatin receptor 2 dimers: A role in receptor trafficking. J Biol Chem 2004;279:36179–36183.
69. Ferré S, Casadó V, Devi LA, Filizola M, Jockers R, Lohse MJ, Milligan G, Pin JP, Guitart X. G protein-coupled receptor oligomerization revisited: functional and pharmacological perspectives. Pharmacol Rev 2014;66:413–434.
70. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol 1981;147:195–197.
71. Levoye A, Dam J, Ayoub MA, Guillaume JL, Couturier C, Delagrange P, Jockers R. The orphan GPR50 receptor specifically inhibits MT1 melatonin receptor function through heterodimerization. Embo J 2006;25:3012–3023.
72. Sharman JL, Benson HE, Pawson AJ, Lukito V, Mpamhanga CP, Bombail V, Davenport AP, Peters JA, Spedding M, Harmar AJ. IUPHAR-DB: updated database content and new features. Nucleic Acids Res 2013;41:D1083–D1088.
73. Jordan BA, Devi LA. G-protein-coupled receptor heterodimerization modulates receptor function. Nature 1999;399:697–700.
74. George SR, Fan T, Xie Z, Tse R, Tam V, Varghese G, O'Dowd BF. Oligomerization of mu- and delta-opioid receptors. Generation of novel functional properties. J Biol Chem 2000;275:26128–26135.
75. Yoshioka K, Saitoh O, Nakata H. Heteromeric association creates a P2Y-like adenosine receptor. Proc Natl Acad Sci U S A 2001;98: 7617–7622.
76. Takahashi K, Furukawa C, Takano A, Ishikawa N, Kato T, Hayama S, Suzuki C, Yasui W, Inai K, Sone S, Ito T, Nishimura H, Tsuchiya E, Nakamura Y, Daigo Y. The neuromedin U-growth hormone secretagogue receptor 1b/neurotensin receptor 1 oncogenic signaling pathway as a therapeutic target for lung cancer. Cancer Res 2006;66: 9408–9419.
77. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L, Wang W, Eigenbrot C, Kaminker JS, Eberhard DA, Waring P, Schuster SC, Modrusan Z, Zhang Z, Stokoe D, de Sauvage FJ, Faham M, Seshagiri S. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 2010;466: 869–873.

78. O'Hayre M, Vázquez-Prado J, Kufareva I, Stawiski EW, Handel TM, Seshagiri S, Gutkind JS. The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. Nat Rev Cancer 2013;13:412–424.
79. Nemoto W, Fukui K, Toh H. GRIP: a server for predicting interfaces for GPCR oligomerization. J Recept Signal Transduct Res 2009;29: 312–327.
80. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V, Conn PJ, Stevens RC. Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. Science 2014;344:58–64.
81. Chattopadhyay A. GPCRs: lipid-dependent membrane receptors. Adv Biol 2014;2014:1–12.
82. Casadó V, Cortés A, Mallol J, Pérez-Capote K, Ferré S, Lluis C, Franco R, Canela EI. GPCR homomers and heteromers: a better choice as targets for drug development than GPCR monomers? Pharmacol Ther 2009;124:248–257.
83. Ganguly S, Clayton AH. a, Chattopadhyay A. Organization of higher-order oligomers of the serotonin 1A receptor explored utilizing homo-FRET in live cells. Biophys J 2011;100:361–368.
84. Paila YD, Kombrabail M, Krishnamoorthy G, Chattopadhyay A. Oligomerization of the serotonin (1A) receptor in live cells: a time-resolved fluorescence anisotropy approach. J Phys Chem B 2011;115: 11439–11447.
85. Prasanna X, Chattopadhyay A, Sengupta D. Cholesterol modulates the dimer interface of the β2-adrenergic receptor via cholesterol occupancy sites. Biophys J 2014;106:1290–1300.
86. Stranahan AM, Cutler RG, Button C, Telljohann R, Mattson MP. Diet-induced elevations in serum cholesterol are associated with alterations in hippocampal lipid metabolism and increased oxidative stress. J Neurochem 2011;118:611–615.
87. Vanmierlo T, Lütjohann D, Mulder M. Brain cholesterol in normal and pathological aging. OCL - Ol Corps Gras Lipides 2011;18:214–217.
88. Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM. Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry 1992;31:12719–12725.
89. Lemmon MA, Flanagan JM, Hunt JF, Adair BD, Bormann BJ, Dempsey CE, Engelman DM. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. J Biol Chem 1992;267:7683–7689.