

MINIREVIEW

The study of G-protein coupled receptor oligomerization with computational modeling and bioinformaticsMarta Filizola¹ and Harel Weinstein^{1,2}

1 Department of Physiology and Biophysics, Weill Medical College of Cornell University, NY, USA

2 Institute for Computational Biomedicine (ICB), Weill Medical College of Cornell University, NY, USA

Keywords

bioinformatics; GPCRs; interface; molecular modeling; oligomerization

Correspondence

M. Filizola, Department of Physiology and Biophysics, Box 75, Weill Medical College of Cornell University, 1300 York Avenue New York, NY 10021, USA

Fax: +01 212 746 8690

Tel: +01 212 746 6348

E-mail: maf2037@med.cornell.edu

(Received 16 February 2005, accepted 8 April 2005)

doi:10.1111/j.1742-4658.2005.04730.x

To achieve a structural context for the analysis of G-protein coupled receptor (GPCR) oligomers, molecular modeling must be used to predict the corresponding interaction interfaces. The task is complicated by the paucity of detailed structural data at atomic resolution, and the large number of possible modes in which the bundles of seven transmembrane (TM) segments of the interacting GPCR monomers can be packed together into dimers and/or higher-order oligomers. Approaches and tools offered by bioinformatics can be used to reduce the complexity of this task and, combined with computational modeling, can serve to yield testable predictions for the structural properties of oligomers. Most of the bioinformatics methods take advantage of the evolutionary relation that exists among GPCRs, as expressed in their sequences and measurable in the common elements of their structural and functional features. These common elements are responsible for the presence of detectable patterns of motifs and correlated mutations evident from the alignment of the sequences of these complex biological systems. The decoding of these patterns in terms of structural and functional determinants can provide indications about the most likely interfaces of dimerization/oligomerization of GPCRs. We review here the main approaches from bioinformatics, enhanced by computational molecular modeling, that have been used to predict likely interfaces of dimerization/oligomerization of GPCRs, and compare results from their application to rhodopsin-like GPCRs. A compilation of the most frequently predicted GPCR oligomerization interfaces points to specific regions of TMs 4–6.

Introduction

The growing experimental evidence showing that GPCR oligomerization has pharmacological and functional implications [1–4] (recent reviews), has prompted the search for detailed structural information of the receptor–receptor interface(s) in order to enable a mechanistic understanding of these complex biological systems. Despite the results from experimental studies

suggesting the participation of C- [5] and N-terminal [6] regions in the self association of rhodopsin-like GPCRs, more recent evidence points to the transmembrane helices (TMs) as the most probable structural elements involved in oligomerization [7–12]. In particular, the recent atomic-force microscopy map of rhodopsin molecules in native mouse disk membranes [10–12] provides direct evidence for the organization of *rhodopsin* protomers into two-dimensional arrays of

Abbreviations

CMA, correlated mutation analysis; GPCR(s), G-protein coupled receptor(s); SCM, subtractive correlated mutation; TM, transmembrane.

dimers. Inference from this map led to the construction of an oligomeric molecular model of rhodopsin with TM4 and TM5 involved in intradimeric contact, while TM1, TM2, and the cytoplasmic loop connecting TM5 and TM6, facilitate the formation of rhodopsin dimer rows.

Two modes of association of the TM helices of GPCR monomers into dimers have been proposed. One of them, which is termed contact dimerization, corresponds to the packing of two different TM bundles with separate binding sites through interactions at interfaces that otherwise would face the lipid environment [13]. The other one, termed 'domain-swapped dimerization', involves interpenetration of transmembrane bundles, where the interacting TMs from two different polypeptides appear as interlaced units [14–17]. Experimental data can be found in the literature to support either contact dimers [10–13,18–21] or domain-swapped dimers [22–29] of GPCRs.

Regardless of the type of geometry assumed for the association of the rhodopsin-like GPCRs, the specific interacting residues that form the dimerization interface remain unknown for most receptor subtypes. To guide specific experiments aimed at producing such information, it is necessary to organize the available data in a structural context. At this stage, only molecular modeling offers such a structural context, and even the most direct experimental evidence for dimer geometry in rhodopsin makes use of such models [11]. The construction of the molecular models of dimers/oligomers for other GPCRs, in the absence of the type of direct data available for rhodopsin, is quite complex. Thus, even for contact dimers in which the seven TM bundles of two GPCR monomers are packed together, there is a large number of alternative possibilities {at least 49 ($= 7 \times 7$) for hetero-dimers and 28 [$= 7(7 + 1)/2$] for homo-dimers}. This number of different alternatives for the possible interfaces can be reduced significantly with the use of a variety of computational methods, including genome level bioinformatics tools and homology models [30,31].

Computational methods designed to help identify protein–protein interaction interfaces (e.g. [32,33]) are especially appropriate for this task. In general, the computational methods that can serve in the modeling of oligomerization interfaces fall into two categories. If 3D structural information is available, approaches generally known as 'docking methods' [33] (review) can be used to identify protein–protein interfaces, based on a variety of selection criteria and exhaustive searches of the 'interaction space'. Still, the accuracy of the predictions from such computational techniques remains quite limited [34]. In contrast, even if structural information

about the interacting G-proteins in a complex is not available, bioinformatics methods based on sequence and genomic information can be used to predict probable regions involved in protein–protein interactions. Many of the limitations and considerations of accuracy and reliability of these methods have been reviewed recently [32].

Some of the computational techniques developed to predict protein–protein interactions have been applied to modeling of GPCR interactions. Most of these approaches utilize sequence and genomic information to predict putative functionally important residues, such as those involved in the interaction of GPCRs with cognate G-proteins [35–41], as well as in signal transduction [35,42–44]. Similarly, bioinformatic methods were tapped for approaches used to predict possible interfaces of GPCR dimerization/oligomerization [16,17,40,45–50]. We review here the underlying principles of these methods, as well as results of their application to rhodopsin-like GPCRs. To enable comparisons, we use the generic numbering system of GPCR sequences (N1.N2) originally described in Balasteros & Weinstein [51]. Specifically, this numbering scheme consists of a number (N1) corresponding to the TM number, and another (N2) counting the position relative to the most conserved residue in the particular TM. The conserved locus is assigned the number 50, and the N2 values of the other loci decrease towards the N-terminus and increase toward the C-terminus.

The compilation of the predicted dimerization/oligomerization interfaces obtained from the comparison shows that regions in TMs 4–6 have the highest number of occurrences. From specific examples, it becomes evident that this type of analysis for GPCR interaction interfaces can provide valuable structure-based hypotheses for further probing of functional mechanisms, because: (a) they can be tested specifically (e.g. with mutagenesis) for sequence determinants of dimerization/oligomerization, as well as for disruption of the dimers and, (b) they provide criteria for the design of experiments to probe the functional effects of dimerization/oligomerization on mechanisms of ligand binding and extent of activation (i.e. pharmacological efficacy), as well as on the involvement of protein–protein interactions in GPCR function (e.g. transactivation).

Methods employed

The evolutionary trace method

The evolutionary trace method is an adaptation of an earlier strategy for the hierarchical analysis of residue

conservation in protein sequence alignments [52]. It was first described by Lichtarge *et al.* [38,53,54] as a technique that predicts functionally important residues (e.g. active sites and functional interfaces) in proteins of known structure. The set of assumptions used to extract such evolutionary trace residues from sequence conservation patterns in homologous proteins includes the following: (a) protein structures descendant from a common ancestor retain their fold (even if sequence identities are as low as 25% [55]), as well as the location of their functional sites and (b) functionally important residues undergo fewer mutations than other residues [56], and their lower mutation rate is interrupted mainly by mutations that cause divergence. Using a dendrogram (or phylogenetic tree) to represent graphically a multiple sequence alignment of homologous proteins, an evolutionary trace residue can be identified as a residue which, upon partitioning of the dendrogram at a certain level of sequence divergence, is conserved within each group into which the dendrogram is divided, but may vary from one group to another.

An example of the steps involved in identifying evolutionary trace residues from a hypothetical multiple sequence alignment is shown in Fig. 1. Based on a dendrogram, sequences in a family can be divided into subfamilies at selected sequence identity cutoffs. At high sequence identity cutoffs, a subfamily is composed of a smaller number of groups of sequences, which are expected to show functional specificity. In contrast, at low sequence identity cutoffs, the subfamilies contain more groups of sequences with less

specificity, but perhaps reflecting functional common elements. In general, the exact cutoff value that must be used to obtain reliable predictions is depending on the protein of interest.

In the example shown in Fig. 1, the dendrogram of a hypothetical multiple sequence alignment was partitioned into four subfamilies using the sequence identity cutoff shown as a vertical dotted line. For each subfamily, a consensus sequence can be compiled by transcribing conserved residues, and leaving variable positions blank, as shown on the right of Fig. 1. Comparison of these consensus sequences allows for the identification of evolutionary trace residues, which are supposed to be part of a protein active site, or a functional interface. More specifically, such residues include: (a) class-specific residues, i.e. residues conserved within a subfamily that are different between subfamilies, but never a gap (e.g. X in Fig. 1) and (b) conserved residues, i.e. residues conserved across the entire sequence family (e.g. E and R in Fig. 1). Residues that cannot be classified as conserved or class-specific residues are called neutral (shown as underscore characters in Fig. 1). Once the evolutionary trace residues have been identified, they can be mapped (e.g. by color-coding) to the structure of one of the proteins in the sequence family, and clustered in 3D space, as shown schematically in Fig. 1.

In the application to over 700 aligned GPCR sequences from classes A (rhodopsin-like), B (secretin-like), and C (metabotropic glutamate-like), an enhanced evolutionary trace method using Monte-Carlo techniques [57] suggested a potential functional site on the lipid-exposed faces of TM5 and TM6 in common to each family or subfamily of receptors [40]. Although this analysis did not result in the identification of exactly the same residues for all the GPCR families and subfamilies studied, the presence of a functional site in the same lipid-exposed region of TM5 and TM6 suggested these helices as candidates for the dimerization interface of GPCRs. As these studies only used the TM regions of GPCRs, the authors could not distinguish between contact and domain-swapped dimers, and suggested that, for the purpose of signaling, the two alternative models are equivalent [16,17,40]. Using the enhanced evolutionary trace method [40], a second functional site on the lipid-exposed faces of TM2 and TM3 was also predicted. Specifically, this functional site was suggested to be implicated either in heterodimerization, or in the formation of higher-order oligomers [17,40]. On the other hand, considerably less functionality was observed on TM1, TM4, and TM7 when using this enhanced evolutionary trace method [40].

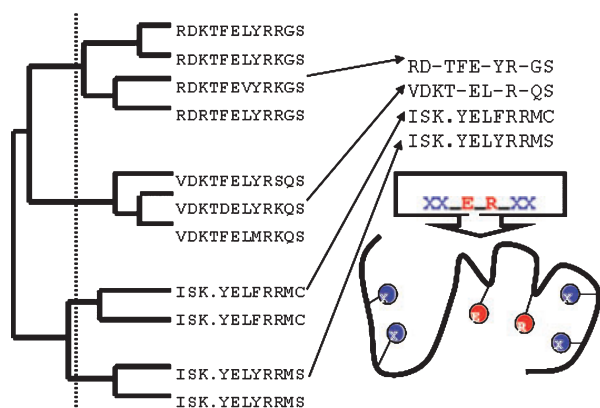


Fig. 1. Steps involved in identifying evolutionary trace residues from a hypothetical multiple sequence alignment. The vertical dotted line identifies the sequence identity cutoff used to partition the dendrogram into four subfamilies. Class-specific and conserved residues, shown in blue and red, respectively, are mapped to the structure of one of the proteins in the sequence family, and clustered in 3D space.

In order to identify clusters of residues that might be responsible for global and class-specific functions, the evolutionary trace method was applied recently to a multiple sequence alignment of visual opsin, bioamine, olfactory, and chemokine class A GPCRs [43]. Among the trace residues suggested to mediate a generic signal transduction mechanism, only one (position 4.47, according to the generic numbering scheme) was predicted to be lipid-exposed based on solvent accessibility values ($> 45\%$) calculated with the GETAREA software version 1.1 [58] using the rhodopsin crystal structure [59]. Interestingly, a mutation at this position in chemokine receptors was recently suggested to affect receptor homodimerization [49].

Correlated mutation analysis

Correlated mutations are typically identified in multiple sequence alignments as loci that mutate simultaneously. A correlated mutation algorithm was described by Gobel *et al.* [60] as a powerful tool to correctly predict physical contacts in homologous proteins. Oliveira *et al.* [61] used a similar approach as Gobel *et al.* [60] to first determine the correlation between residue positions in GPCRs. Such inferences based on correlated mutation analysis (CMA) have been verified experimentally to indicate spatial adjacencies between GPCR TMs in the intramolecular portion of the proteins (e.g. [62,63]). In addition, the observation that the type of compensatory changes identified by CMA tend to accumulate at protein interfaces [61,64] led to the extension of the concept of correlated mutations to predict protein–protein contacts. This approach is based on the reasoning that sequence changes that occur during evolution at one of the interaction interfaces must be compensated by changes in the other interacting protein in order to preserve the protein–protein interface. This

implementation is illustrated schematically in Fig. 2. Specifically, sequence changes that occur during evolution at the interaction interface of a given protein (Fig. 2B) must be compensated by changes in the interacting protein (Fig. 2C) to preserve the protein–protein interface. In the hypothetical multiple sequence alignment shown in Fig. 2, these compensatory changes occur at positions 36 and 140. This basic principle was used not only to identify likely interfaces of GPCR–G-protein interactions [35–37,41,43,45], but also to predict interfaces of homo- and heterodimerization in GPCRs [16,17,30,31,45,47–49].

Correlated mutation analysis performed by Gouldson *et al.* [45] on multiple sequence alignments of bioaminergic, somatostatin, neurokinin, opioid, thyrotrophin, and chemokine receptors, using the WHATIF molecular graphics software [65], showed an accumulation of correlated mutations on the lipid-exposed surface of these receptors. Specifically, the occurrence of correlated mutations on the external faces of TM1, TM5, TM6, and TM7 was interpreted as an indication that these helices may be involved in the formation of domain-swapped dimers. In contrast, conformational changes and/or the formation of higher order structures were invoked to explain the simultaneous appearance of correlated mutations on the external faces of TM2, TM3, and particularly TM4.

This type of approach was enhanced with filtering algorithms to enable identification of the likely hetero- and homo-oligomerization interfaces of family A GPCRs [47,48]. These methods take advantage of an improved CMA-based algorithm [66] that combines correlated mutations with other types of sequence properties (e.g. sequence conservation and contact density), and utilize the structural information from the rhodopsin crystal structure [59] as the basis to predict functionally important residues at the dimerization

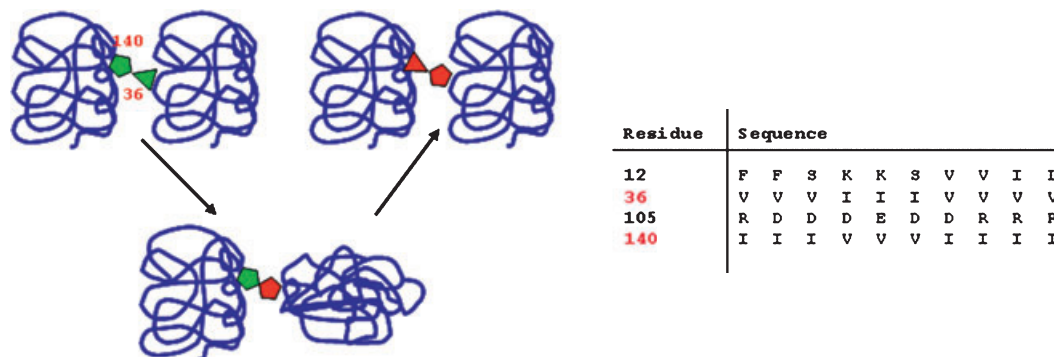


Fig. 2. Schematic representation of the concept of correlated mutations applied to protein–protein interactions. In the specific example, compensatory changes occur at positions 36 and 140 of the hypothetical multiple sequence alignment.

interfaces of GPCRs. Specifically, the approach we developed to identify probable interfaces of GPCR heterodimerization, termed Subtractive Correlated Mutation (SCM) method [47], consists of a modified version of the original algorithm [66] that provides a means to filter out the intramolecular pairs of correlated residues within each interacting monomer from the complete list of intra- and intermolecular pairs of correlated residues. These correlated mutations are identified in a multiple sequence alignment of concatenated monomeric sequences of the two different GPCRs, obtained from the same organisms. The concatenation step is essential for the identification of the heterodimerization interface, as outlined in detail [47]. A similar approach was developed independently by Pazos & Valencia [67].

Although a powerful bioinformatics tool that was demonstrated from specific tests to identify residues that are functionally essential, the CMA in itself does not usually achieve specific identification of the residue composition of the dimerization interfaces of GPCRs (e.g. results of Gouldson *et al.* 2001, which essentially show correlated mutations in all seven TM helices). Consequently, additional stringency criteria must be added to the CMA approaches, especially in the application to GPCR homo-dimerization, in order to achieve reliable predictions of the dimerization/oligomerization interface of GPCRs. To this end, we introduced the filtering criteria described below, which are applied to the list of calculated correlated mutations to reduce the number of false positives [31], and are combined with the geometric filtering derived from the construction of 3D molecular models of putative configurations of GPCR dimers/oligomers [48]. Experience shows that application of these criteria may cause the procedure to overlook some correct predictions of residues involved in the actual interface, i.e. false negatives. Nevertheless, this multistep filtering remains advantageous because it ensures elimination of isolated residues (as opposed to complete interfaces), which have high correlation index values. Briefly, we developed an approach in which the correlated pairs of residues are first sorted by decreasing correlation values (from 1 to 0). To increase the chance of obtaining correctly predicted contacts, only highly correlated pairs are taken into account for each case. This is achieved by limiting the predictions to a maximum of $L/2$, where L is the length of the sequence of each receptor subtype (i.e. only the most significant $L/2$ predictions are retained, and the others discarded). This specific filtering is carried out because a list of $L/2$ was demonstrated to be enriched in correctly predicted contacts [66]. However, even the $L/2$ list is purged of correlated

pairs with a correlation index ≤ 0.7 , in order to reduce the number of false positives. Conversely, however, if the number of correlated pairs with a correlation index equal to 1 exceeds the number of $L/2$ correlated pairs, they are all taken into account. Finally, the location of the identified residues on the 'outward' facing portions of the monomers is established using the information contained in the crystal structure of the cognate rhodopsin receptor [59]. Any pair of correlated residues with either one, or both inaccessible to solvent (thus not considered to be outward-facing) is eliminated from the list of predictions. Specifically, only pairs of correlated residues where both positions have a surface exposure of more than 45 \AA^2 are considered as candidates for intermolecular contacts (the implicit assumption is that association of GPCRs occurs only via contact dimers/oligomers). A final filtering criterion is based on the requirement for interaction neighborhoods on the proposed helix interface. Specifically, one definition of the neighborhood that we have used [47,48] is that residues in a TM helix are considered to define an interface only if at least three appear close to each other, within $i + 7$. Application of this final filter further reduces the number of false positives obtained from the bioinformatics approach of CMA.

In a recent application to rhodopsin-like GPCR subtypes for which homo-dimerization has been demonstrated experimentally [31], our enhanced CMA-based approach identified TM1 and TM4 most often as putative interfaces among the studied GPCRs. The frequency with which these two TMs appear in the predictions suggested them as the most probable segments of rhodopsin-like GPCRs to be involved in dimerization/oligomerization interfaces. This finding is intriguing given the recent experimental data suggesting a role for precisely TM1 and TM4 in the dimerization/oligomerization of rhodopsin-like GPCRs (including rhodopsin [11], dopamine D2 [9], $\alpha 1$ adrenergic [7], and C5a [68] receptors).

Methods that detect tree-determinant positions

Two of the fully automatic methods recently implemented in Valencia's lab to detect tree-determinant positions in multiple sequence alignments [69] (specifically, the Level Entropy and the SequenceSpace Automation methods) were combined recently with CMA [60,67] to identify probable interfaces of dimerization in chemokine receptors [49]. Tree-determinant positions correspond to residues that are conserved within a subfamily of proteins, but differ between subfamilies. Several algorithms have been developed to search for the best tree-determinants involved in the function of a

protein family [52,53,70–79]. It is clear that each of the different implementations of this bioinformatics approach presents special advantages as well as specific drawbacks. However, comparing the results of all these methods is beyond the scope of this review.

The main features of the approach are illustrated here for the combination of the Level Entropy method and the SequenceSpace Automatization method with CMA. The Level Entropy method searches automatically for different partitions of the phylogenetic tree of a protein family in order to identify an optimal partition according to the number of tree-determinants involved in the function of the protein family, normalized by the number of conserved positions in each subfamily. The distance between the distribution of tree-determinants and the product of the distributions of conserved positions in each subfamily is calculated using the concept of Relative Entropy from Information Theory [80]. In some protein families there could be more than one optimal dendrogram partition that features tree-determinants involved in different functions, but this method has been shown to identify the most informative partition according to the number of tree-determinants involved in biological activity [69].

The SequenceSpace Automatization Method consists of an automated version of the earlier SequenceSpace analysis method [70], which had been shown to work more effectively than other approaches in the prediction of functionally important residues [81]. This new automatic implementation of the method attempts to reduce human intervention in the recognition of residues with similar tendencies by using a geometric criterion that identifies clusters of residues in the multi-dimensional space. Both the Level Entropy method and the SequenceSpace Automatization Method have been tested on nonredundant lists of protein families, and demonstrated to predict residues that have a clear tendency to be close to functionally important residues.

A combination of these two methods together with CMA [60,67] was applied to the chemokine receptor family to predict a specific interface of dimerization in the chemokine receptor corresponding to the SWISS-PROT sequence identity code *CKR5_HUMAN* [49]. Specifically, TM1, TM2, and TM4 were proposed as candidates of the homodimerization interface of chemokine receptors. The predicted helices were then used as a guide to build 3D models of the *CKR5_HUMAN* homodimer using the automated docking procedure embedded into the GLOBAL RANGE MOLECULAR MATCHING docking program [82]. The potential homodimer model was selected from among 100 initial solutions proposed by this docking program. Using the correct

membrane orientation and the proximity of the calculated lipid-exposed tree-determinants and/or correlated residues as criteria for selection, this proposed homodimer model produced an asymmetric interface involving TM1 and TM4 helices. Interestingly, experimental evaluation of several mutants with alterations in TM1 and TM4 identified a two-point mutation I52V/V150A at positions 1.54 and 4.47, respectively, as responsible for the disruption of the homodimer of *CKR5_HUMAN* [49].

Hidden-site class model of evolution

As shown for the methods described above, most of the bioinformatics techniques that detect putative functional sites in proteins from multiple sequence alignments, share the assumption that proteins that are evolutionarily related might exhibit common structural and functional features corresponding to detectable patterns in their sequences. As a result, a suitable representation of the evolutionary relationships between proteins under study is an essential requirement for the prediction of sequence locations bearing structural or functional significance. In general, evolutionary relationships between proteins can be represented by a matrix indicating the rate at which every amino acid substitution occurs during evolution. Current models of evolution use a single substitution matrix for all locations in all protein sequences. This is, however, a limitation of these models because the probability that an amino acid substitution at a particular location in the sequence of a protein would produce any functional effect is not the same at all locations. A novel model of evolution, termed hidden-site class model [83–85], was proposed to overcome this limitation by using different substitution matrices to represent amino acid substitutions at different locations in a protein sequence. Specifically, each location in a multiple sequence alignment can be described by one of the different types of sites, thus creating site classes that are each associated with a specific substitution model. While the assignment of locations to different site classes is unknown a priori, it can be calculated iteratively after optimization of the corresponding substitution models using a maximum likelihood formulation. This hidden-site class method was demonstrated to attain better phylogenetic inferences by identifying locations in the protein sequences that are considered to be under similar selective pressure, and by characterizing changes in this selective pressure. Specifically, locations that are assigned to site classes with the slowest rate of substitution are expected to correspond to structural or functional important positions.

In an application to 199 aminergic receptors from the class A family of GPCRs [50], this hidden-site class model of evolution identified 56 locations that belonged to the slowest evolving site classes in one of the site class models expected to provide the most reasonable insights into the structural and functional features of the aminergic receptors. Among them, 16 of 33 locations that are known to be involved in ligand binding in aminergic receptors were identified. The method also detected lipid-exposed evolutionarily conserved locations on TM4, TM5 and TM6 in different subfamilies. Specifically, a general abundance of lipid-exposed locations on TM5 and TM6 of most aminergic receptors was interpreted to suggest the involvement of these helices in the dimerization of the aminergic receptors, whereas TM4 and TM5 were suggested to be involved in the dimerization of muscarinic, opsin, and serotonin receptors.

Common elements in the prediction of GPCR–GPCR interaction interfaces

The various computational studies that looked for possible dimerization/oligomerization interfaces of GPCRs, using the type of bioinformatics approaches

described above, did not predict exactly the same interfaces for all the GPCR subfamilies studied. This is not entirely surprising given the differences not only in methodology, but also in the selected data sets and the corresponding multiple sequence alignments. Both the assumptions underlying the computational algorithms and the selection of the sequences in the alignment determine the nature of the answers returned by the application of the computational tools. The statistical nature of these tools makes their success in predicting likely dimerization/oligomerization interfaces of GPCRs strongly dependent on the number of sequences available for each family or subfamily of these proteins. As more sequences become available with the completion of sequencing of more genomes, the power of these approaches is expected to increase significantly.

In spite of the specific differences in the results from different methods, it is notable that some TM segments appear more often than others in the prediction of GPCR interfaces. To zoom into the level of specific lipid-exposed residues, we mined the prediction data (Fig. 3) – to see whether some loci were more frequently predicted to be at interfaces than others – by comparing the results of the different studies

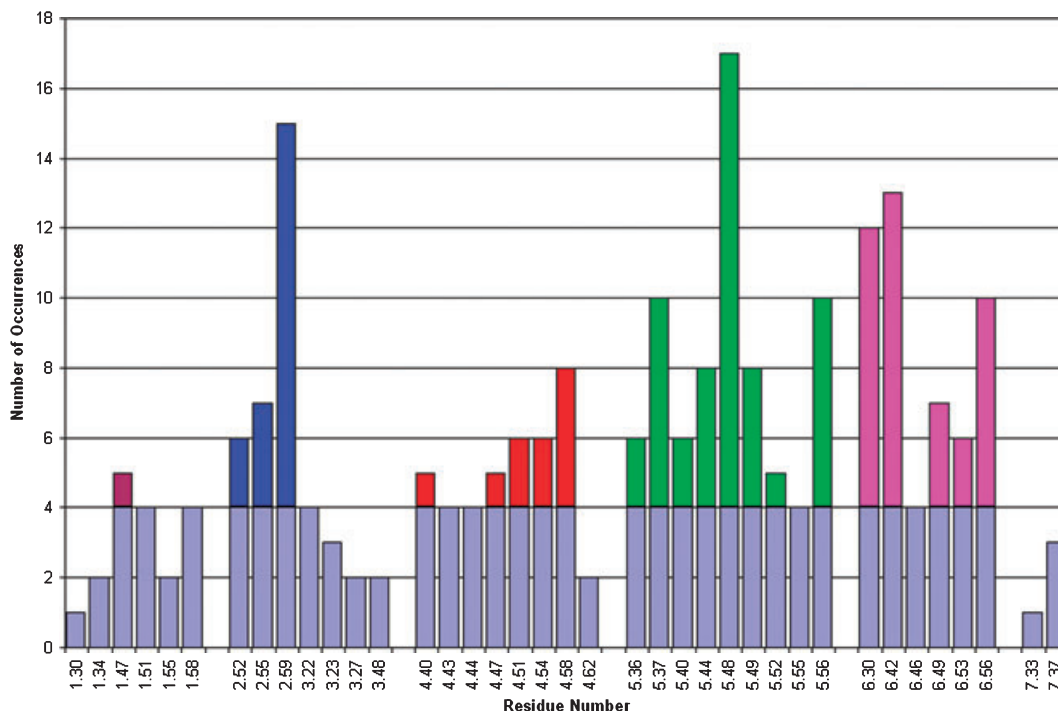


Fig. 3. Occurrence of the lipid-exposed residues in the predictions from bioinformatics methods applied to search for dimerization/oligomerization interfaces of GPCRs. Residues within a TM region that are predicted more than a baseline of four times are indicated for each TM (purple, blue, red, green and magenta for residues in TM1, TM2, TM4, TM5 and TM6, respectively).

[16,17,30,31,40,45,47,48,50] that have looked so far for possible dimerization/oligomerization interfaces of rhodopsin-like GPCRs. Among all predicted residues in each study, we only compared the lipid-exposed residues within TM regions, and subject to the solvent accessibility criterion of values $> 45\%$. These values were calculated with the GETAREA software version 1.1 [58] using the atomic coordinates of the rhodopsin crystal structure [59]. It is important to keep in mind, however, that rhodopsin and rhodopsin-like GPCRs may exhibit structural differences that affect local exposure of residues to the environment [86]. This can create differences in the details of GPCR interfaces, and therefore inaccurate estimation of solvent accessibility values assigned to equivalent positions in different receptors.

Figure 3 presents a histogram plot of the number of times that each lipid-exposed residue has been predicted with the methods discussed here to belong to dimerization/oligomerization interfaces of GPCRs. Residues within a TM region that were predicted more often than a base line of four times, are indicated for each TM (purple, blue, red, green, and magenta for residues in TM1, TM2, TM4, TM5 and TM6, respectively). Most of these residues are within TM4, TM5, and TM6, indicating that the prediction of dimerization/oligomerization interfaces of GPCRs with various computational methods has thus far pointed to a specific role for the lipid-exposed regions of these three helices. Among the loci identified within each of these three helices, 4.58, 5.48 and 6.42 have the greatest number of occurrences. In particular, the latter has almost the same frequency of occurrence as 6.30. (Note that the large frequency of occurrence for 6.30, at the boundary between TM6 and the cytoplasmic loop connecting TM5 and TM6, could be explained by an involvement of this locus in a broader oligomerization scheme of GPCRs. This is suggested by the atomic force microscopy map of rhodopsin in native membranes [11], which indicates that the cytoplasmic loop connecting TM5 and TM6 facilitates the formation of rows of rhodopsin-dimers.)

Interestingly, position 4.58, which corresponds to a cysteine in dopamine D_2 receptor, was shown recently to form a copper phenanthroline-induced disulfide cross-link resulting in the appearance of a dimeric band [9]. This finding is consistent with the hypothesis that TM4 is involved in a symmetrical interface in dopamine D_2 receptor dimers, and that 4.58 is part of this interface.

To the best of our knowledge, no information about the involvement of position 5.48 in the dimerization/oligomerization of rhodopsin-like GPCRs exists

in the literature. In contrast, a phenylalanine at this position in the 5HT2 subfamily of serotonin receptors has been suggested to be involved in ligand binding [87]. On the other hand, early experimental studies pointed to the involvement of 6.42 based on inhibition assays with synthetic peptides comprising the amino acid sequence of TM6 of β_2 -adrenergic receptor that contains the glycoporphin-like dimerization motif GXXXG [88]. Thus, the involvement of position 6.42 (G280 in human β_2 -adrenergic receptor) in the dimerization/oligomerization of GPCRs was inferred. However, these data do not necessarily establish TM6 as the dimer interface in β_2 -adrenergic receptor, because a specific peptide-receptor interaction at one site may modulate the ability of the receptor to form dimers at a different interface.

Last but not least, it is important to emphasize here that the specific lipid-exposed positions shown in Fig. 3 may also have been picked out by the bioinformatics tools due to functional roles different from dimerization/oligomerization. Such functional roles may include correct protein folding, required interactions with the lipid bilayer, or interactions with other proteins. The high number of occurrences of position 2.59 seems to relate to this hypothesis because the proline residue found at this position in almost all aminergic receptors, on which most of the studies focused, may induce a functional kink in this helix [86] that could affect the structural integrity of the receptors.

The specific reason for the structure/function role of the predictions shown in Fig. 3 notwithstanding, they probably represent valuable hypotheses for further experimental exploration of functionally important residues located on the surface of rhodopsin-like GPCRs. Interpreted in the structural context offered by models of the GPCRs [30,86,89] and their oligomers [11,48], the putative interfaces are ripe for specific probing with dimerization-disrupting mutations, cross-linking, and derivatization. Such experiments should yield valuable insight not only about the structural details of the GPCR-GPCR interaction in both homo- and heteromeric complexes, but will offer specific tools for the elucidation of the functional roles of GPCR interactions and the molecular details of signaling.

Structural details from computational modeling and bioinformatics

As anticipated above, structural details from computational modeling and bioinformatics may provide valuable hypotheses for the experimental exploration of interfaces of GPCR oligomerization. The goal is to identify key residues responsible for disrupting the

GPCR oligomeric structures in order to probe effects on receptor function, and achieve a better understanding of the mechanisms that regulate these complex systems.

Such a process of inquiry is illustrated by an approach developed to identify the molecular determinants for the oligomerization of rhodopsin-like GPCRs as an iterative protocol of computational prediction and experimental validation [31]. In this protocol, results from CMA-based methods are used to guide the construction of 3D molecular models of GPCR homo- and heteromers [47,48]. Inferences from these models (and not from correlated mutations alone) are then implemented in the design of key collaborative experiments that serve to probe, validate, and refine the hypotheses of interaction between monomers (e.g. an important validation probe is the design of dimerization-disrupting mutants). Specifically, the experimental designs can use a cysteine cross-linking method that was recently implemented to identify residues at the homo-dimerization interfaces of dopamine D₂ receptors. The success of the application of this experimental procedure is documented in the literature [9].

An illustration of the structural details that were identified for dopamine D₂ receptor dimers using this combined computational and experimental strategy is given in Fig. 4. Our CMA-based approach for the identification of homodimerization interfaces of GPCRs [48] predicted that 4.58, 4.48, 4.55, and 4.60

are part of the interface of dopamine D₂ receptor dimers. These predictions were then used to build geometrically feasible configurations of dopamine D₂ receptor homodimers using the criteria reported in [48], as reviewed above. Notably, the initial (unrefined) models suggested that position 4.60, which is located at the extracellular boundary of TM4, should not be considered in the construction of the dimer models, because of its possible inaccurate solvent accessibility value (loop regions are missing in these models). Figure 4 shows the three residues that were used to guide the construction of the receptor homodimeric interface in CPK representation. Among them, C4.58 at the extracellular end of TM4 had already been proposed from cysteine cross-linking experiments to be part of a symmetrical interface in dopamine D₂ receptor dimers [9]. Additional cysteine cross-linking experiments (with copper-phenanthroline) were then carried out at the predicted positions 4.48 and 4.55, and these positions were also suggested to be at the interface between interacting monomers of the dopamine D₂ receptor (W. Guo, L. Shi, M. Filizola, H. Weinstein, J. A. Javich, unpublished results). The refined 3D models that incorporate the new information are currently being used to refine the scheme of dimerization disruption by single mutations (or a stretch identifiable as a motif) of dopamine D₂ receptor homodimers. The predictive ability of this combined computational and experimental strategy in the application to dopamine

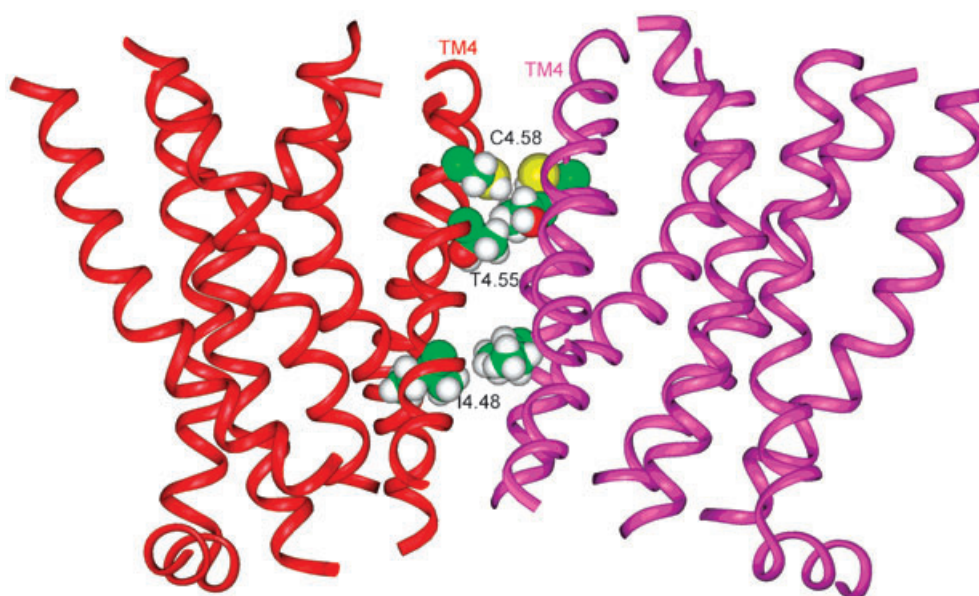


Fig. 4. Proposed interface for dopamine D₂ receptor homodimers. Specific residues predicted with our CMA-based method to be at the interface between monomers are shown in CPK representations.

receptors creates exciting expectations for the study of the structural and functional properties of oligomerization in other GPCRs as well, and possibly also in any other membrane protein. Moreover, it points to the exciting possibilities for exploration of functional mechanisms of GPCRs in a structural context offered by models of the protein–protein interaction. The types of structure-based hypotheses that can be derived from such models are essential for the development of mechanistic understanding of the multibody, time-dependent elements of GPCR function in cellular-signaling pathways.

Experimentally tested predictions

Most experimental methods that have been used so far to study GPCR oligomerization, such as inhibition assays using synthetic peptides, coexpression, bioluminescence resonance energy transfer (BRET), and fluorescence resonance energy transfer (FRET), do not reveal the details of the GPCR oligomerization interface. In particular, they do not reveal the specific residues that are in actual contact. As a result, only a few positions have been reported in the literature to participate directly to the oligomerization interface(s) of GPCRs based entirely on experimental data. Specifically, these positions are: 1.54 and 4.47 in the chemokine receptors [49]; 4.58 in the dopamine D₂ receptor [9] and 6.42 in the β_2 -adrenergic receptor [88]. Notably, three of these four positions had been predicted with bioinformatics approaches to lie at the dimerization interface more often than the four times considered as a base line (see Fig. 3).

As detailed information about the oligomerization interface of GPCRs is essential for designing dimerization-disrupting mutants, and thus for probing structural and mechanistic hypotheses by interfering with GPCR function, more incisive experimental approaches are needed that are better suited for the identification of specific residues at the dimeric/oligomeric interface of GPCRs. The close collaboration established between the computational and experimental approaches to investigation of GPCR oligomerization, as illustrated in this review, should then enable rapid success in unraveling the mechanisms and functional implications of GPCR oligomerization in GPCR signaling processes.

Acknowledgements

We are grateful to our colleagues Drs Wen Guo, Jonathan Javitch, and Lei Shi for illuminating collaborative studies, and to Dr Masha Niv for comments on the manuscript. Computational support was provided

by the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputing Center. The authors also acknowledge access to the computer and bioinformatics facilities at the Institute of Computational Biomedicine (ICB) of Weill Medical College. The work was supported in part by NIH grants P01 DA12923, and K05 DA-00060.

References

- Breitwieser GE (2004) G protein-coupled receptor oligomerization: implications for G protein activation and cell signaling. *Circ Res* **94**, 17–27.
- Bai M (2004) Dimerization of G-protein-coupled receptors: roles in signal transduction. *Cell Signal* **16**, 175–186.
- Terrillon S & Bouvier M (2004) Roles of G-protein-coupled receptor dimerization. *EMBO Report* **5**, 30–34.
- Milligan G (2004) G protein-coupled receptor dimerization: function and ligand pharmacology. *Mol Pharmacol* **66**, 1–7.
- Cvejic S & Devi LA (1997) Dimerization of the delta opioid receptor: implication for a role in receptor internalization. *J Biol Chem* **272**, 26959–26964.
- AbdAlla S, Zaki E, Lothar H & Quitterer U (1999) Involvement of the amino terminus of the B (2) receptor in agonist-induced receptor dimerization. *J Biol Chem* **274**, 26079–26084.
- Carrillo JJ, Pediani J & Milligan G (2003) Dimers of class A G protein-coupled receptors function via agonist-mediated trans-activation of associated G proteins. *J Biol Chem* **278**, 42578–42587.
- Carrillo JJ, Lopez-Gimenez JF & Milligan G (2004) Multiple interactions between transmembrane helices generate the oligomeric {alpha}1b-adrenoceptor. *Mol Pharmacol* **66**, 1123–1137.
- Guo W, Shi L & Javitch JA (2003) The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer. *J Biol Chem* **278**, 4385–4388.
- Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A & Palczewski K (2003) Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* **421**, 127–128.
- Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K & Engel A (2003) Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem* **278**, 21655–21662.
- Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A & Palczewski K (2004) The G protein-coupled receptor rhodopsin in the native membrane. *FEBS Lett* **564**, 281–288.
- Schulz A, Grosse R, Schultz G, Gudermann T & Schöneberg T (2000) Structural implication for receptor oligomerization from functional reconstitution studies of

- mutant V2 vasopressin receptors. *J Biol Chem* **275**, 2381–2389.
- 14 Gouldson PR & Reynolds CA (1997) Simulations on dimeric peptides: evidence for domain swapping in G-protein-coupled receptors? *Biochem Soc Trans* **25**, 1066–1071.
 - 15 Gouldson PR, Snell CR & Reynolds CA (1997) A new approach to docking in the beta 2-adrenergic receptor that exploits the domain structure of G-protein-coupled receptors. *J Med Chem* **40**, 3871–3886.
 - 16 Gouldson PR, Snell CR, Bywater RP, Higgs C & Reynolds CA (1998) Domain swapping in G-protein coupled receptor dimers. *Prot Eng* **11**, 1181–1193.
 - 17 Gouldson PR, Higgs C, Smith RE, Dean MK, Gkoutos GV & Reynolds CA (2000) Dimerization and domain swapping in G-protein-coupled receptors: a computational study. *Neuropsychopharmacology* **23**, S60–S77.
 - 18 Lee SP, O'Dowd BF, Ng GY, Varghese G, Akil H, Mansour A, Nguyen T & George SR (2000) Inhibition of cell surface expression by mutant receptors demonstrates that D2 dopamine receptors exist as oligomers in the cell. *Mol Pharmacol* **58**, 120–128.
 - 19 Hamdan FF, Ward SD, Siddiqui NA, Bloodworth LM & Wess J (2002) Use of an in situ disulfide cross-linking strategy to map proximities between amino acid residues in transmembrane domains I and VII of the M3 muscarinic acetylcholine receptor. *Biochemistry* **41**, 7647–7658.
 - 20 Hadac EM, Ji Z, Pinon DI, Henne RM, Lybrand TP & Miller LJ (1999) A peptide agonist acts by occupation of a monomeric G protein-coupled receptor: dual sites of covalent attachment to domains near TM1 and TM7 of the same molecule make biologically significant domain-swapped dimerization unlikely. *J Med Chem* **42**, 2105–2111.
 - 21 Overton MC & Blumer KJ (2002) The extracellular N-terminal domain and transmembrane domains 1 and 2 mediate oligomerization of a yeast G protein-coupled receptor. *J Biol Chem* **277**, 41463–41472.
 - 22 Maggio R, Vogel Z & Wess J (1993) Coexpression studies with mutant muscarinic/adrenergic receptors provide evidence for intermolecular 'cross-talk' between G-protein-linked receptors. *Proc Natl Acad Sci USA* **90**, 3103–3107.
 - 23 Ridge KD, Lee SS *et al.* (1996) Examining rhodopsin folding and assembly through expression of polypeptide fragments. *J Biol Chem* **271**, 7860–7867.
 - 24 Kobilka BK, Kobilka TS, Daniel K, Regan JW, Caron MG & Lefkowitz RJ (1988) Chimeric $\alpha 2$ -, $\beta 2$ -adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science* **240**, 1310–1316.
 - 25 Schoneberg T, Liu J & Wess J (1995) Plasma membrane localization and functional rescue of truncated forms of a G protein-coupled receptor. *J Biol Chem* **270**, 18000–18006.
 - 26 Schoneberg T, Yun J, Wenkert D & Wess J (1996) Functional rescue of mutant V2 vasopressin receptors causing nephrogenic diabetes insipidus by a co-expressed receptor polypeptide. *EMBO J* **15**, 1283–1291.
 - 27 Gudermann T, Schoneberg T & Schultz G (1997) Functional and structural complexity of signal transduction via G-protein-coupled receptors. *Annu Rev Neurosci* **20**, 399–427.
 - 28 Nielsen SM, Elling CE & Schwartz TW (1998) Split-receptors in the tachykinin neurokinin-1 system – mutational analysis of intracellular loop 3. *Eur J Biochem* **251**, 217–226.
 - 29 Bakker RA, Dees G, Carrillo JJ, Booth RG, Lopez-Gimenez JF, Milligan G, Strange PG & Leurs R (2004) Domain swapping in the human histamine h1 receptor. *J Pharmacol Exp Ther* **311**, 131–138.
 - 30 Filizola M, Visiers I, Skrabanek L, Campagne F & Weinstein H (2003) Functional mechanisms of GPCRs in a structural context. In *Strategies in molecular neuropharmacology* (Schousboe A & Brauner-Osborne H, eds), pp. 235–266. Humana Press, Totowa, NJ.
 - 31 Filizola M, Guo W, Javitch JA & Weinstein H (2005) Oligomerization domains of G-protein coupled receptors: insights into the structural basis of GPCR association. In *Contemporary Clinical Neuroscience: the G-protein Coupled Receptor Handbook* (Devi LA, ed), pp. 243–265. Humana Press, Totowa, NJ.
 - 32 Valencia A & Pazos F (2002) Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol* **12**, 368–373.
 - 33 Smith GR & Sternberg MJ (2002) Prediction of protein–protein interactions by docking methods. *Curr Opin Struct Biol* **12**, 28–35.
 - 34 Dixon JS (1997) Evaluation of the CASP2 docking section. *Proteins Supplement* **1**, 198–204.
 - 35 Oliveira L, Paiva AC & Vriend G (2002) Correlated mutation analyses on very large sequence families. *Chembiochem* **3**, 1010–1017.
 - 36 Oliveira L, Paiva AC & Vriend G (1999) A low resolution model for the interaction of G proteins with G protein-coupled receptors. *Protein Eng* **12**, 1087–1095.
 - 37 Horn F, van der Wenden EM, Oliveira L, AP IJ & Vriend G (2000) Receptors coupling to G proteins: is there a signal behind the sequence? *Proteins* **41**, 448–459.
 - 38 Lichtarge O, Bourne HR & Cohen FE (1996) Evolutionarily conserved Galphabeta gamma binding surfaces support a model of the G protein-receptor complex. *Proc Natl Acad Sci USA* **93**, 7507–7511.
 - 39 Lichtarge O, Sowa ME & Philipp A (2002) Evolutionary traces of functional surfaces along G protein signaling pathway. *Methods Enzymol* **344**, 536–556.
 - 40 Dean MK, Higgs C, Smith RE, Bywater RP, Snell CR, Scott PD, Upton GJ, Howe TJ & Reynolds CA (2001)

- Dimerization of G-protein-coupled receptors. *J Med Chem* **44**, 4595–4614.
- 41 Moller S, Vilo J & Croning MD (2001) Prediction of the coupling specificity of G protein coupled receptors to their G proteins. *Bioinformatics* **17**, S174–S181.
- 42 Oliveira L, Paiva PB, Paiva AC & Vriend G (2003) Sequence analysis reveals how G protein-coupled receptors transduce the signal to the G protein. *Proteins* **52**, 553–560.
- 43 Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG & Lichtarge O (2004) Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *J Biol Chem* **279**, 8126–8132.
- 44 Suel GM, Lockless SW, Wall MA & Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* **10**, 59–69.
- 45 Gouldson PR, Dean MK, Snell CR, Bywater RP, Gkoutos G & Reynolds CA (2001) Lipid-facing correlated mutations and dimerization in G-protein coupled receptors. *Prot Eng* **14**, 759–767.
- 46 Gkoutos GV, Higgs C, Bywater RP, Gouldson PR & Reynolds CA (1999) Evidence for dimerization in the β_2 -adrenergic receptor from the evolutionary trace method. *Intl J Quantum Chem Biophys* **Q74**, 371–379.
- 47 Filizola M, Olmea O & Weinstein H (2002) Prediction of heterodimerization interfaces of G-protein coupled receptors with a new subtractive correlated mutation method. *Prot Eng* **15**, 881–885.
- 48 Filizola M & Weinstein H (2002) Structural models for dimerization of G-Protein coupled receptors: the opioid receptor homodimers. *Biopolymers (Peptide Sci)* **66**, 317–325.
- 49 Hernandez-Falcon P, Rodriguez-Frade JM, Serrano A, Juan D, del Sol A, Soriano SF, Roncal F, Gomez L, Valencia A, Martinez AC & Mellado M (2004) Identification of amino acid residues crucial for chemokine receptor dimerization. *Nat Immunol* **5**, 216–223.
- 50 Soyer OS, Dimmic MW, Neubig RR & Goldstein RA (2003) Dimerization in aminergic G-protein-coupled receptors: application of a hidden-site class model of evolution. *Biochemistry* **42**, 14522–14531.
- 51 Ballesteros JA & Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci* **25**, 366–428.
- 52 Livingstone CD & Barton GJ (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci* **9**, 745–756.
- 53 Lichtarge O, Bourne HR & Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* **257**, 342–358.
- 54 Lichtarge O, Yamamoto KR & Cohen FE (1997) Identification of functional surfaces of the zinc binding domains of intracellular receptors. *J Mol Biol* **274**, 325–337.
- 55 Chothia C & Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *Embo J* **5**, 823–826.
- 56 Zvelebil MJ, Barton GJ, Taylor WR & Sternberg MJ (1987) Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J Mol Biol* **195**, 957–961.
- 57 Upton G & Fingleton B (1985) *Spatial Data Analysis by Example*. Wiley, Chichester.
- 58 Fraczekiewicz R & Braun W (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Computational Chem* **19**, 319–333.
- 59 Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, LeTrong I, Teller DC, Okada T, Stenkamp RE *et al.* (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**, 739–745.
- 60 Gobel U, Sander C, Schneider R & Valencia A (1994) Correlated mutations and residue contacts in proteins. *Proteins* **18**, 309–317.
- 61 Oliveira L, Paiva ACM & Vriend G (1993) A common motif in G-protein coupled seven transmembrane helix receptors. *J Comp Aid Mol Des* **7**, 649–658.
- 62 Zhou W, Flanagan C, Ballesteros JA, Konvicka K, Davidson JS, Weinstein H, Millar RP & Sealfon SC (1994) A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Mol Pharmacol* **45**, 165–170.
- 63 Sealfon SC, Chi L, Ebersole BJ, Rodic V, Zhang D, Ballesteros JA & Weinstein H (1995) Related contribution of specific helix 2 and 7 residues to conformational activation of the serotonin 5-HT_{2A} receptor. *J Biol Chem* **270**, 16683–16688.
- 64 Pazos F, Helmer-Citterich M, Ausiello G & Valencia A (1997) Correlated mutations contain information about protein–protein interaction. *J Mol Biol* **271**, 511–523.
- 65 Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8** (52–6), 29.
- 66 Olmea O & Valencia A (1997) Improving contact predictions by the combination of correlated mutations and other sources of sequence information. *Fold Des* **2**, S25–S32.
- 67 Pazos F & Valencia A (2002) *In silico* two-hybrid system for the selection of physically interacting protein pairs. *Proteins* **47**, 219–227.
- 68 Klco JM, Lassere TB & Baranski TJ (2003) C5a receptor oligomerization. I. Disulfide trapping reveals oligomers and potential contact surfaces in a G protein-coupled receptor. *J Biol Chem* **278**, 35345–35353.

- 69 del Sol Mesa A, Pazos F & Valencia A (2003) Automatic methods for predicting functionally important residues. *J Mol Biol* **326**, 1289–1302.
- 70 Casari G, Sander C & Valencia A (1995) A method to predict functional residues in proteins. *Nat Struct Biol* **2**, 171–178.
- 71 Dorit RL & Ayala FJ (1995) ADH evolution and the phylogenetic footprint. *J Mol Evol* **40**, 658–662.
- 72 Andrade MA, Casari G, Sander C & Valencia A (1997) Classification of protein families and detection of the determinant residues with an improved self-organizing map. *Biol Cybern* **76**, 441–450.
- 73 Zhang B, Rychlewski L, Pawlowski K, Fetrow JS, Skolnick J & Godzik A (1999) From fold predictions to function predictions: automation of functional site conservation analysis for functional genome predictions. *Protein Sci* **8**, 1104–1115.
- 74 Goh CS, Bogan AA, Joachimiak M, Walther D & Cohen FE (2000) Co-evolution of proteins with their interaction partners. *J Mol Biol* **299**, 283–293.
- 75 Landgraf R, Xenarios I & Eisenberg D (2001) Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. *J Mol Biol* **307**, 1487–1502.
- 76 Armon A, Graur D & Ben-Tal N (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J Mol Biol* **307**, 447–463.
- 77 Madabushi S, Yao H, Marsh M, Kristensen DM, Philippi A, Sowa ME & Lichtarge O (2002) Structural clusters of evolutionary trace residues are statistically significant and common in proteins. *J Mol Biol* **316**, 139–154.
- 78 Lichtarge O & Sowa ME (2002) Evolutionary predictions of binding surfaces and interactions. *Curr Opin Struct Biol* **12**, 21–27.
- 79 Pupko T, Bell RE, Mayrose I, Glaser F & Ben-Tal N (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* **18** (Suppl. 1), S71–S77.
- 80 Shannon C & Weaver W (1963) *Mathematical Theory of Communication*. University of Illinois press, Champaign, IL.
- 81 Pazos F, Sanchez-Pulido L, Garcia-Ranea JA, Andrade MA, Atrian S & Valencia A (1997) Comparative analysis of different methods for the detection of specific regions in protein families. In *Biocomputing and Emergent Computation* (Lundh, D, Olsson, B & Narayanan, A, eds), pp. 132–145. World Scientific, Singapore.
- 82 Vakser IA & Jiang S (2002) Strategies for modeling the interactions of transmembrane helices of G protein-coupled receptors by geometric complementarity using the GRAMM computer algorithm. *Methods Enzymol* **343**, 313–328.
- 83 Koshi JM & Goldstein RA (1995) Context-dependent optimal substitution matrices. *Protein Eng* **8**, 641–645.
- 84 Koshi JM & Goldstein RA (1998) Models of natural mutations including site heterogeneity. *Proteins* **32**, 289–295.
- 85 Koshi JM, Mindell DP & Goldstein RA (1999) Using physical-chemistry-based substitution models in phylogenetic analyses of HIV-1 subtypes. *Mol Biol Evol* **16**, 173–179.
- 86 Ballesteros JA, Shi L & Javitch JA (2001) 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* **60**, 1–19.
- 87 Shapiro DA, Kristiansen K, Kroeze WK & Roth BL (2000) Differential modes of agonist binding to 5-hydroxytryptamine (2A) serotonin receptors revealed by mutation and molecular modeling of conserved residues in transmembrane region 5. *Mol Pharmacol* **58**, 877–886.
- 88 Hebert TE, Moffett S, Morello JP, Loisel TP, Bichet DG, Barret C & Bouvier M (1996) A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J Biol Chem* **271**, 16384–16392.
- 89 Visiers I, Ballesteros JA & Weinstein H (2002) Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Methods Enzymol* **343**, 329–371.