Prediction and targeting of GPCR oligomer interfaces

Carlos A.V. Barreto[a], Salete J. Baptista[a,b], António José Preto[a], Pedro Matos-Filipe[a], Joana Mourão[a,c], Rita Melo[a,b], Irina Moreira[a,d,*]

[a]Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal
[b]Centro de Ciências e Tecnologias Nucleares, Instituto Superior Técnico, Universidade de Lisboa, CTN, LRS, Portugal
[c]Institute for Interdisciplinary Research, University of Coimbra, Coimbra, Portugal
[d]Science and Technology Faculty, University of Coimbra, Coimbra, Portugal
[*]Corresponding author: e-mail address: irina.moreira@cnc.uc.pt

Contents

1. Introduction 2
2. Characterization and prediction of oligomer interfaces 5
   2.1 Experimental approaches 5
   2.2 Computational approaches 7
   2.3 Machine learning approaches 14
3. Targetting PPIs: Orthosteric and allosteric modulation 28
   3.1 Orthosteric modulation 28
   3.2 Allosteric modulation 30
4. Concluding remarks 30
Acknowledgments 31
References 32

Abstract

GPCR oligomerization has emerged as a hot topic in the GPCR field in the last years. Receptors that are part of these oligomers can influence each other's function, although it is not yet entirely understood how these interactions work. The existence of such a highly complex network of interactions between GPCRs generates the possibility of alternative targets for new therapeutic approaches.

However, challenges still exist in the characterization of these complexes, especially at the interface level. Different experimental approaches, such as FRET or BRET, are usually combined to study GPCR oligomer interactions. Computational methods have been applied as a useful tool for retrieving information from GPCR sequences and the few X-ray-resolved oligomeric structures that are accessible, as well as for predicting new and trustworthy GPCR oligomeric interfaces.

Machine-learning (ML) approaches have recently helped with some hindrances of other methods. By joining and evaluating multiple structure-, sequence- and co-evolution-based features on the same algorithm, it is possible to dilute the issues

Progress in Molecular Biology and Translational Science
ISSN 1877-1173
https://doi.org/10.1016/bs.pmbts.2019.11.007

of particular structures and residues that arise from the experimental methodology into all-encompassing algorithms capable of accurately predict GPCR-GPCR interfaces.

All these methods used as a single or a combined approach provide useful information about GPCR oligomerization and its role in GPCR function and dynamics. Altogether, we present experimental, computational and machine-learning methods used to study oligomers interfaces, as well as strategies that have been used to target these dynamic complexes.

1. Introduction

G protein-coupled receptors (GPCRs) superfamily has been a subject of high interest in cell and molecular biology field for decades, mainly due to its presence in various physiological events. However, their mechanism of action is not yet fully understood. This family also represents 34% of the drugs approved by the Food and Drug Administration (FDA), demonstrating its high viability as therapeutic targets.[1] GPCRs are highly dynamical proteins that mediate the signal transduction triggered by extracellular stimuli through the cell membrane. The family has up to 800 different receptors divided by their structural and functional similarities into 5 major subfamilies: class A, B, C, frizzled, and adhesion. GPCRs have a common structure present through the different subfamilies: seven transmembrane domains (TM), connected by three extracellular (ECL) and three intracellular loops (ICL), with N-terminal in the extracellular side and the C-terminal on the intracellular side. The TM region is highly conserved, and interhelical bonds and hydrophobic interactions maintain its stability. Loops are the least conserved regions and display structural variability between the subfamilies.[2,3]

For many years the GPCR family members have been studied as monomeric entities; however, in recent years, accumulating evidence has shown that GPCRs can function in dimeric (homo and hetero) or higher-order oligomeric states. Class C GPCRs are known to form dimers constitutively through their extensive extracellular domain to work.[4] Class A has increasingly data pointing toward the existence of homo and heterodimers.[5,6] The ratio between monomeric and dimeric states is a defining characteristic of this subfamily. There is evidence that suggests that in receptors like B_2 adrenergic receptor (B_2AR) and α_{1B}-adrenergic receptor ($\alpha_{1B}AR$), dimerization is necessary for efficient surface localization.[7,8] Thus, some authors suggested that dimers are assembled during biosynthesis, perhaps inside the endoplasmic reticulum.[9–11] Nevertheless, recent data proposes a dynamical view of GPCR dimers that are in equilibrium with their monomeric

forms and have variable timescales, depending on the membrane or cellular environments (such as cytoskeleton and scaffolding or anchoring proteins).[9,12–15]

The effects of dimerization/oligomerization on the structure and dynamics of receptors are not yet entirely understood, neither their implication in human physiology and pathology. However, the steady increase in studies related to the allosteric interactions between the receptors in complex have brought light into this subject. A simple way to view the importance of these interactions is to categorize them in three groups as it was done by Guidolin et al.[16]: (a) neighbor receptors can modulate each other's orthosteric binding site; (b) receptors can modulate the intracellular binding pocket, thus altering signaling pathways; (c) or new allosteric sites can emerge for binding with different modulators.

There are currently several curated and specialized databases where information concerning 3D structures of GPCRs can be found, including dimers, and other membrane proteins (MPs). Some of these databases include: (a) the MPs of known 3D structure (mpstruc)[17] that identifies and collect MPs of the PDB data bank (as of October 7, 2019 it contains 952 unique entries); (b) the Transporter Classification DataBase (TCDB)[18] that provides functional and phylogenetic information on membrane transport proteins (as of October 7, 2019 it contains ~1405 families of transport proteins); (c) the Protein Data Bank of Transmembrane Proteins (PDBTM)[19] that uses TMDET algorithm[20] in all PDB entries for location of TM protein in the lipid bilayer (as of October 7, 2019 it contains 4084 transmembrane proteins); (d) the Orientations of Proteins in Membrane (OPM) database[21] that uses PPM server to provide spatial arrangements of MPs with respect to the hydrocarbon core of the lipid bilayer; (e) the MemProtMD, a meta-database that presents the results of molecular dynamics simulations of some MPs of mpstruc embedded in lipid bilayers (the database contains ~3500 intrinsic MPs structures)[22]; and (f) more specific databases for GPCRs such as the G-Protein Coupled Receptor Database (GPCRdb) with 15,147 proteins (as of October 7, 2019),[23] the G-Protein Coupled Receptor Oligomerization Knowledge Base (GPCR-OKB),[24] or the GPCR-HGmod[25] that contains 1026 putative 3D structural models of GPCRs in the human genome generated by the GPCR-I-TASSER pipeline and deposited in the GPCR-EXP (database of experimentally solved and predicted GPCR structures) (https://zhanglab.ccmb.med.umich.edu/GPCR-EXP/). Known GPCR-GPCR interactions are stored and can be acquired through GPCR-OKB[24] and GPCR-HetNet.[26]

Currently, there are 12 structures of GPCR dimers in PDB that present a crystallographic asymmetric unit and with a software-determined quaternary structure[27] (PDB id: 2VT4,[28] 4GPO,[29] 3ODU,[30] 3OE9,[30] 4EA3,[31] 6AK3,[32] 5O9H,[33] 5ZKQ,[34] 3CAP,[35] 2PED,[36] 2J4Y,[37] 4JKV,[38] 6N52[39]). Furthermore, three additional structures are found as an asymmetric unit but with no quaternary structure prediction: two from class A (PDB id: 5UEN,[40] 4DJH[41]) and one from class C (PDB id: 2E4U[42]).

The existence of this highly complex network of interactions between GPCRs and how they can modulate each other's behavior contributed to the development of new therapeutic approaches. Nevertheless, the challenges in characterizing these complexes remain, and in particular at the interface level, which plays a unique role in the development of new targeting drugs. Herein, we present experimental, *in silico* computational methods and ML methods that are currently in use for the characterization and interpretation of these interfaces. We also review which strategies have been used to target these dynamic complexes. Some essential key concepts for further understanding of the chapter are presented in Box 1.

BOX 1 Key concepts

Artificial neural network (ANN): is a machine learning (ML) tool with a graph-based architecture inspired in the brain and how neurons connect and interact with each other. ANNs comprise a series of nodes (also called neurons) where mathematical operations are performed on the features fed to the system.[118] Nodes are connected by edges with associated weights and biases, updated according to the overall performance of the ANN (the system can learn from those changes in weights' values).[119] Nodes in ANNs are organized in layers (the input, output and hidden layers). Data is fed to the system through the input layer, and the response to the problem (value or class) is given through the output layer. Hidden layers connect the input and output layers. Networks comprising more than one hidden layer are considered Deep Neural Networks (DNNs).[120]

Random forest (RF): is an ensemble model of decision trees.[121]

Ensemble systems: is a ML algorithm that output a model that can gather the contributions of several models. The individual models contribute according to voting systems, that can vary in name, such as hard or soft voting, depending on the grade of contribution for the overall result.[122]

Hot spots (HS): HS residues are defined as those that, upon alanine mutation, generate a binding free energy difference ($\Delta\Delta G_{binding}$) $\geq$2.0 kcal/mol.[123] These residues are more prone to be structural and functionally relevant, increasing their influence in binding processes.

BOX 1 Key concepts—cont'd

Paralogs genes: type of homologous genes that arise by gene-duplication events from the last common ancestor. In these events, functional gene novelty is free to change during evolution.[124]

Machine learning: is a subfield of Artificial Intelligence that gives to the computer, by algorithms, the ability to learn a pattern from a large amount of heterogeneous data. The prediction of the best possible solution is reached by training the algorithm using a training set and scoring its performance using a validation set.[125] The model is finally trained using an independent test set. ML models for interface prediction usually use supervised learning (in which the output is known) with regression or classification (e.g., classifying a surface residue as interfacial or non-interfacial) algorithms.

Protein interaction interface: non-uniform surface areas between two protein monomers that allow more energetically favorable interactions to occur. They are characterized by two main regions, the "core" and the "rim," that are different in terms of physicochemical properties and evolutionary conservation.[126] Interfaces can be located in obligate (protomers are not stable structures *in vivo* unless they are in a complex) and transient (binding partners may dissociate from each other and exist as stable entities in the unbound state *in vivo*) complexes interactions.

Protein surface: the exterior hydrophilic environment where polar residues reside mainly. Interfacial and non-interfacial residues are part of the protein surface.

Solvent-accessible surface area (SASA): is the atomic surface area of a molecule accessible to a solvent (usually expressed in square Ångstroms, $Å^2$).[127]

Support vector machine (SVM): are ML models on which the various examples in the dataset are represented in *n* dimensions hyperspace (where n is the number of features describing the sample).[128] The model tries to find hyperplanes dividing the various examples according to the classes to which they belong to, generating regions that define each class.[129]

Transmembrane protein: is a type of integral polytopic protein that crosses the entire cell membrane and stays permanently attached to it.

2. Characterization and prediction of oligomer interfaces

2.1 Experimental approaches

Experimental-based methods can be applied to study protein-protein interactions (PPIs), including GPCR oligomers. These approaches can be split

into four categories, affinity-based methods, proteomics-based methods, fluorescence-based assays and genetic assays. Schiedel et al.[43] performed an extensive review about the application of experimental methods to investigate GPCR oligomers. To study PPIs in GPCR oligomerization, different experimental approaches are usually combined, being the most common the fluorescent-based assays in combination with affinity-based methods and/or genetic assays. Most of the PPIs described by Schiedel et al.,[43] which were taken from GPCR oligomer complexes, were identified by using two or more experimental techniques. Noteworthy, several parameters should be considered when choosing the most suitable method, such as the nature of the interaction (permanent or transient; weak or strong), and the determination of complex stoichiometry.[43]

Fluorescence-based assays are the most applied approaches to study GPCR oligomerization. It includes frequently used methods such as FRET (fluorescence resonance energy transfer), BRET (bioluminescence resonance energy transfer), BiFC (biomolecular fluorescence complementation assays), and most recently time-resolved FRET (Tr-FRET). However, these methodologies are mainly used to confirm the GPCR oligomerization, rather than to provide information concerning the size of oligomeric structures or even their inherent dynamic nature.[43] FRET, for instance, is unable to detect PPIs that specifically occur at the membrane.[44] More recent microscopy-based approaches, such as total internal reflection fluorescence microscopy (TIRFM) and single-molecule imaging have been applied to give insights into dynamics of GPCR oligomers and to supply information about oligomerization state of GPCRs.[43,45]

A new AlphaScreen-Based Assay was recently identified and applied to study GPCR oligomerization, namely the Dopamine receptor D_2 (D_2R)/adenosine A_{2A} receptor ($A_{2A}R$) heterodimerization, confirming for the first time the existence of this heterodimer in human caudate nucleous.[46] Moreover, a combined approach using proximity ligation assay and co-immunoprecipitation experiments was used to disclose the first evidence of Bradykinin B_2 receptors (B2R).[47] Additionally, a new FRET-based strategy focused on ligand binding selectivity for oligomers was also identified.[48] This approach, which is easy to implement and adaptable for high-throughput screening, allows the identification of hetero-oligomer specific ligands. It can also be used as a starting point to disclose insights into oligomer crosstalks, ultimately leading to the expose of new and critical features that can be of high interest for the treatment of several diseases associated with GPCR oligomerization.

2.2 Computational approaches

Although the number of crystal structures has been steadly increasing in the last years, a limited number of oligomer GPCR structures are available.[49,50] Moreover, the experimental determination of GPCR oligomer structures at a detailed molecular level is still a difficult challenge.[45] Therefore, computational methods have been applied as a useful tool for retrieving information about the few X-ray-resolved GPCR oligomeric structures that are accessible, as well for predicting new and trustworthy GPCR oligomeric interfaces.[27,51] Computational approaches to study GPCR dimerization or higher-order oligomers can typically be divided into two groups: sequence-based and structured-based methodologies.[52]

2.2.1 Sequence-based methods

Sequence-based methods take into consideration the protein sequence to predict residues engaged in the dimerization interface.[53] Depending on sequence conservation, this branch of computer-assisted approaches is subdivided into two classes: (i) assuming a evolutionary conservation of interface dimers among all proteins of a subfamily and (ii) postulating that the dimerization interface can change among members of the same subfamily, during the evolutionary process. Evolutionary Trace (ET) method,[54] Correlated Mutation Analysis (CMA),[55] and Subtractive Correlated Mutation (SCM) are examples of the first class, whereas Differential Evolutionary Trace (DET), Spatial Cluster Detection (SCD) and Hidden-Site Class Model of Evolution are included in the second one.[45,56]

Evolutionary Trace Method takes into account that proteins of the same family evolving from a common ancestor should have a similar backbone structure, and as such should have a higher degree of conservation in sites that contribute to its function, hence displaying a lower mutation frequency at these positions.[54,57] This method was applied for studying putative GPCR dimerization interfaces by using Monte Carlo techniques. The multiple alignments of >700 GPCR sequences revealed an important potential functional site on the lipid-exposed faces of TM5 and TM6. This study allowed the identification of a second putative functional site on TM2 and TM3, which, besides the previous one, suggests that GPCRs can oligomerize. The ability to dimerize is common among GPCR family or subfamily, rather than a specific feature of GPCR members.[58]

Correlated Mutation Analysis is a method that searches for pairs of residues in a multiple sequence alignment (MSA) that remain conserved or are

mutated together during evolution, suggesting that the effect of a mutation is compensated by another one to keep protein-protein interface functional.[55,59,60] This technique has been applied to predict MP interfaces, namely on GPCR. A correlated mutation analysis applied to a group of class A GPCRs showed that a significant number of correlated mutations are allocated on the external region (lipid-exposed) of the helices, proposing that these regions can constitute sites for PPIs. Therefore, the correlated residues can be involved in important conformational changes in the receptor, as well as in the formation of GPCR homodimers or heterodimers.[61]

Subtractive Correlated Mutation method corresponds to an improved version of CMA by applying filtering algorithms capable of discarding the intramolecular pairs of correlated residues in both monomers analyzed from all described correlated residue pairs. Thus, the residues that constitute the dimer interface could be identified. The application of this methodology allowed the identification of TM4, TM5, and TM6 of the δ-opioid receptor (DOR) and TM1 of the μ-opioid receptor (MOR) as critical in the formation of heterodimer interfaces, since the correlated residues were found on the external (lipid-facing) surface of those TMs.[60]

Differential Evolutionary Trace was first applied by Madabushi et al.[62] to study the GPCR family. The sequence alignment of different subfamilies of class A GPCRs was performed to identify residues ultimately responsible for global and class-specific activities. This method filtered out the trace residues among all sequences from the ones among the target subtype, thus determining the remaining residues as displaying a specific role for the subtype analyzed.[56,62] Among the ET residues identified by this approach, only one appears to be involved in the homodimerization of CCR5, with some controversial studies about its importance.[56]

Hidden-Site Class Model of Evolution method applies different matrices to represent amino acid substitutions at diverse locations in a protein sequence, overcoming the limitation introduced by the majority of models used to study PPIs involving evolutionary relationships. Those models use a single substitution matrix for all locations in all sequences, which could lead to inaccurate predictions, as the likelihood of an amino acid substitution at a specific location in the protein sequence will not necessarily promote the same functional effect at all locations.[63–67] Hidden-Site Class Model was applied to GPCR to perform a family-specific analysis (the study included 199 class A GPCRs, one of the most studied classes). The results of this study highlighted the presence of lipid-facing evolutionary conserved locations on TM5 and TM6 for the majority of aminergic families, as well as on TM4 and

TM5 for muscarinic and opsin families.[68] Altogether, this approach led to an improvement in the prediction of functionally important residues.[45]

Spatial Cluster Detection (SCD) approach was first developed to predict GPCR oligomeric interfaces by analyzing the spatial distribution of conserved residues on the molecular surface of a specific GPCR subtype. Unlike Hidden-Site Class Model, this method takes into account both the subtype interface specificity and the assessment of the spatial location of the detected residues. Moreover, it was based on the assumption that oligomeric GPCR interfaces are located on the molecular surface of each monomer, and the interface residues are conserved within the same GPCR subtype. This methodology was applied to different GPCRs subtypes to test its performance. The different studies showed that the predicted interfaces of β_2 adrenergic receptor, D_2R, and rhodopsin were in agreement with the experimentally determined interfaces, even though the oligomeric interface region was different among the three GPCR subtypes under investigation.[69–73] GRIP server, a tool for GPCR oligomeric interface prediction, is based on SCD.[74]

Unlike structure-based approaches, the information provided by sequence-based techniques about oligomer interface-forming residues cannot be translated into 3D dimer structures.[75] This fact could explain why the data retrieved from sequence-based approaches to study GPCR oligomerization is more relevant when in conjugation with structure-based methods, rather than when used as a single approach.[50]

2.2.2 Structure-based methods

Structure-based methodologies play an essential role in the prediction of GPCR oligomeric interfaces, as well as in the study of its interaction dynamics.[43] Protein-protein docking and molecular dynamics, MD (Classic and Coarsed-Grain) are among the main structure-based approaches applied in the prediction of GPCR oligomeric interfaces.

2.2.2.1 Protein-protein docking

Protein-protein docking is a widely used method that takes advantage of being faster and less costly in terms of time and computational resources than other structure-based techniques applied in oligomerization prediction, namely MD simulations.[75] Although the majority of protein-protein docking studies were based on a rigid-body approach, the most current approaches take into account, leastwise, receptor side-chain flexibility.[43] However, most applied docking approaches used to investigate PPIs were

generally developed for water-soluble proteins, which makes them inadequate for GPCR oligomeric prediction. In fact, many available protein-protein docking softwares consider parameters that are optimized for soluble proteins, such as desolvation energy, which is an inaccurate criterion to be taken into account for modeling MPs.[51] In an attempt to investigate which protein-protein docking tool available is the most suitable for study transmembrane proteins, namely GPCR oligomer interfaces, Kaczor et al.[76] compared eight protein-protein docking softwares: ClusPro, GRAMM-X, HADDOCK, HEX, PatchDock, SymmDock, and ZDOCK. By analyzing different multimeric transmembrane proteins retrieved from PDB, they found that GRAMM-X software, which includes an evolutionary conservation term in its scoring function, granted the best docking results. Moreover, this study also showed that the protein-protein docking tools under investigation were able to predict transmembrane protein complexes, which display a larger interface and are rich in cavities. That fact could justify the unsuccessful results obtained in the prediction of GPCR dimeric interfaces by applying the available protein-protein docking approaches.[76]

In recent years, some protein-protein docking softwares, such as DOCK/PIERR[77] and Rosetta MP,[78] developed a specific version for modeling MPs. A specific protein-protein docking algorithm, Memdock,[79] was also developed for α-helical transmembrane proteins, showing improved docking accuracy in comparison with standard protein-protein docking algorithms. Another specific protein-protein docking-based protocol able to accurately predict GPCR dimer interfaces was developed by using protein-protein docking with Rosetta software and external scoring. This approach was validated against a series of GPCR dimers, and the obtained results were mostly in line with experimental and modeling data.[80] This multi-component protocol was applied to generate a model of D_2R homodimer in an inactive conformation, which was further used to investigate the interaction of different bivalent antagonists with that receptor. The results obtained revealed an asymmetric dimer model with the TM4-TM5-TM7-TM1 interface as the best-scored model.[81]

Recently, HADDOCK 2.1 was applied in combination with experimental studies (BRET) and molecular dynamics simulations (MD) to develop a new approach able to map GPCR dimer interfaces. The structural model of A_2AR-D_2R heterodimer with TM4/5 interface was generated, giving insights into the structural basis that underlies allosteric modulation, thus constituting a further step in the development of drugs acting on central nervous system disorders. This multi-approach can also be used as a starting

point to study the interface of many GPCR hetero-complexes.[82] In fact, protein-protein docking applied to GPCR modeling is often used as a multi-step approach that includes experimental data and MD simulations to be validated. This multi-approach is necessary due to the lack of accuracy of protein-protein docking, namely the limited ability to consider protein flexibility, which can be overcome by applying MD simulations.[83]

2.2.2.2 Molecular dynamics

MD simulations have been used to study GPCR dimers for several years, and the topic has been extensively reviewed by Altwaijry et al.,[84] Simpson et al.,[85] Selent and Kaczor,[86] and Guo et al.[45] This method provides a higher spatial resolution over a longer timescale than any other computational approach. As a structure-based method, the MD protocol starts with a required 3D-structure of the dimer, which can be obtained from (i) a crystal structure available, (ii) a homology-based model, or (iii) a docking-based model. In classical MD, simulations involving GPCR dimer or higher-order oligomers are performed with fully atomistic conditions using CHARMM[87–89] and/or AMBER[90] forcefields. However, due to the high computational costs of all-atom simulations, the timescale applied is between nano- and microseconds, which can only reveal small conformational changes.

To solve some limitations of the classical MD, coarse-grained MD (CGMD) has been extensively used in this field. In this type of simulations, the fully atomistic structure of the dimer/oligomer is converted into a simplified version where small beads represent residues. Therefore, multiple replicas with extended timesteps can be retrieved by using CGMD.[84] This is important because a single run of MD cannot describe the properties of the system accurately due to their random Gaussian behavior. An ensemble of independent replicas with different initial conditions is needed to have an accurate representation of the system.[27] The Martini[91] forcefield is commonly used in this type of simulation.

Classical MD and CGMD have been extensively used to study GPCR oligomerization, in particular how the dynamics of the receptor are altered by oligomerization and to predict interfaces. A review of the literature available applying MD/CGMD to study GPCR oligomers is presented in Table 1. The analysis of results from MD/CGMD should always consider any available experimental data, although it is challenging to correlate experimental snapshots with the different potential states of a GPCR oligomer.[27]

Table 1 Summary of the MD studies on GPCR dimers available on the literature.

Type	System	Method	Forcefield	Reference
Homodimer	Rho/Rho	MD	GROMOS87	92
		MD	OPLSAA	93
		MD	Amber/parm99	94
		CGMD	Martini	84,95,96
		CGMD + MD	Martini + CHARMM36	97
	β_2AR/β_2AR	CGMD	Martini	98–100
		CGMD with umbrella sampling	Martini	101
	β_1AR/β_1AR	CGMD	Martini	84,100
		CGMD with umbrella sampling	Martini	101
		CGMD + MD	Martini + CHARMM36	97
	CXCR4/CXCR4	MD	OPLSAA	102
		CGMD	Martini	84
		CGMD	Martini	103
		CGMD + MD	Martini + CHARMM36	97
	δOR/δOR	CGMD	Martini	104
		CGMD with umbrella sampling	Martini	105,106
		CGMD + MD	Martini + CHARMM36	97
	μOR[inactive]/μOR[inactive]	CGMD	Martini	104
	μOR[inactive]/μOR[inactive]	CGMD + MD	Martini + CHARMM36	97,107
	μOR[inactive]/μOR[inactive]	Unbiased CGMD + biased CGMD	Martini	108
	μOR[inactive]/μOR[active]	CGMD + MD	Martini + CHARMM36	107

	μOR[active]/μOR[active]	Unbiased CGMD + biased CGMD	Martini	
	κOR/κOR	CGMD	Martini	104
		CGMD + MD	Martini + CHARMM36	97
	$A_{2A}R/A_{2A}R$	CGMD + MD	Martini + CHARMM36	97
	A_3R/A_3R	MD	Amber7 FF	109
	D_2R/D_2R	MD	OPLSAA	81
	LHR/LHR	MD	CHARMM	110
	mGluR1/mGluR1	CGMD + MD	Martini + CHARMM36	97
	NTS1/NTS1	CGMD	Martini	111
	AT_1R[inactive]/AT_1R[inactive]	MD	CHARMM36	112
	AT_1R[inactive]/AT_1R[active]	MD	CHARMM36	112
	AT_1R[active]/AT_1R[active]	MD	CHARMM36	112
Hetereodimer	δOR[inactive]/μOR[inactive]	MD	GROMOS87	113
		CGMD	Martini	104
		CGMD + MD	Martini + CHARMM36	114
	δOR[inactive]/μOR[active]	CGMD + MD	Martini + CHARMM36	114
	δOR/κOR	CGMD	Martini	104
	$A_1R/A_{2A}R$ [in complex with Gi and Gs]	MD	AMBER99SB	115
	$A_{2A}R/D_2R$	MD	OPLSAA	82
	mGluR2/5-HT_{2A}	MD	CHARMM 22/27	116
Homotetramer	$(V_2R)_4$	MD	CHARMM 22/27	117

2.3 Machine learning approaches

In the previous sections, we reviewed the characterization and prediction of oligomerization interaction interfaces in GPCRs, using experimental methods (*in vitro* or *in vivo*) and computational tools (*in silico*). Here, in this section, we focus on ML predictive methods of interaction interfaces occurring in homo- and hetero-oligomeric MPs, such as GPCRs (some essential key concepts for further understanding of the chapter are presented in Box 1 and Fig. 1). Jones and Thornton[131] wrote an extensive review on this subject. According to the authors, the interface between two protein chains can be characterized according to six major dimensions: (i) size and shape; (ii) electrostatic complementarity; (iii) residue interface propensities; (iv) hydrophobicity; (v) secondary structure; and (vi) complex formation. However, the definition of protein residues as interfacial has proven to be difficult using those measures since no agreement was found when settling the cut-off values. Recently, other criteria have been proposed for the definition of an interface,[132] simplifying and clarifying the concept, enhancing reproducibility and allowing this kind of data to be

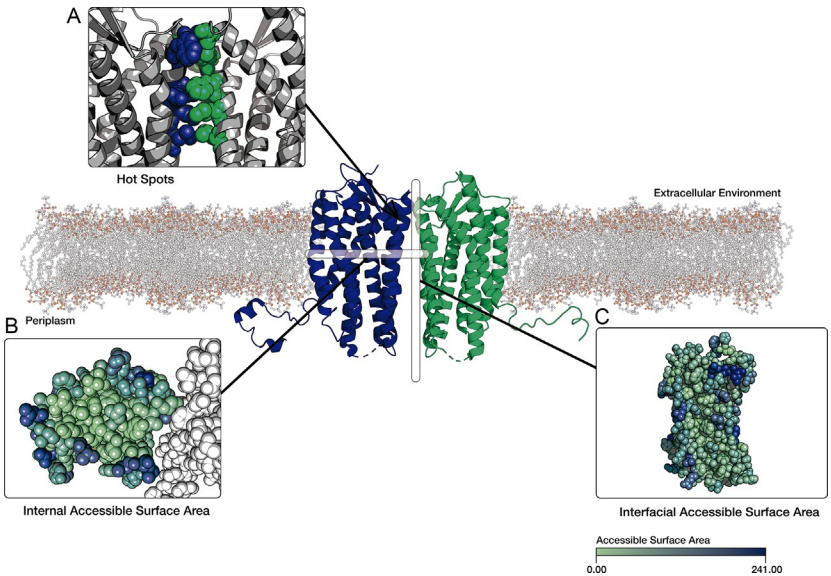

Fig. 1 Structure of the CXCR4 chemokine receptor dimer (transmembrane helices of chain A are displayed in blue and of chain B in green—PDB id: 3ODU)[30] inserted into a membrane bilayer (PDB id: 2MLR[130]). Additionally, three insight windows are presented: (A) hot-spot region present within the interchain interface; (B and C) SASA of all residues from a longitudinal (internal) and interfacial point of view, respectively.

handled by using automated tools, such as ML. Hence, bear in mind that using the same definition is a critical aspect when comparing and evaluating different prediction methods currently available in the literature. Thus, an interface can be defined based on:

(i) ΔASA *(variation in accessible surface area)*: a particular residue is classified as interfacial if a variation in its ASA upon complexation ($\Delta ASA = {}_{complex}ASA - {}_{monomer}ASA$) is larger than 1 $Å^2$.

(ii) *Heavy atom distance*: a residue is considered interfacial if any heavy atom (non-hydrogen atom) between two interacting protein chains is within a determined threshold diameter, usually ranging from 4 to 6 Å.[133]

(iii) $C_\alpha - C_\alpha$ *distance*: two residues in different chains interact if their C_α atoms are within a determined distance from each other (Xue et al. suggested a distance of 8 Å).[132]

(iv) *van der Waals surface distance*: two residues located in different chains display interactions if their van der Waals surfaces are within a distance of approximately 0.5 Å from each other.[134]

In contrast with MPs, to date, several ML algorithms have been used to predict interaction sites and interfaces of soluble proteins, such as Naïve Bayes,[135] artificial neural networks (ANNs),[136] support vector machines (SVMs)[137] or random forests (RFs)[138] (others examples are in Table 2). Although these approaches could also be used for GPCRs and other MPs, the modulation of PPIs interfaces by the hydrophobic environment of the membrane hinders this application.[99] Besides that, challenges in extraction and crystallization of proteins have limited the number of solved TM proteins three-dimensional (3D) structures deposited in public databases that can be further used to extract information of PPI interface residues for ML models. This lack of structures is particularly noticeable for those proteins involved in transient binding interactions where the number of experimental determined 3D structures is even lower.[132]

A common approach of ML models used to predict PPIs and interfaces in new proteins is based on binary classification problems that used databases containing experimental determined interacting residues (see Table 3 for some examples of protein-protein interactions and interface databases) to train the algorithms.

Nowadays, data-driven models try to exploit a combination of unique characteristics of interfacial residues from both computational and experimental methods to understand the nature of the intermolecular interactions and to improve model performance (for more details see Tuncbag et al.[182]).

Table 2 Alphabetically ordered list of ML-based methods and web-servers for prediction and identification of PPIs.

Methods name	Type of ML algorithm	Main features	Server or meta server URL	Type of model	Reference
BIPSPI	XGBoost	BIPSPI was trained with sequence- (amino acid type, sequence profiles, conservation scores) and structural-based features (geometrical descriptors, hydrophobicity, secondary structure, half-sphere exposure and contact number) from both protein partners of each complex. The method was developed to a partner-specific prediction of residue-residue contacts and binding sites	http://bipspi.cnb.csic.es	Sequence- and structural-based	139
ComplexContact	DL	DL method that makes use of co-evolution information, sequential features and contact occurrence patterns	http://raptorx2.uchicago.edu/ComplexContact	Sequence and co-evolution based	140
cons-PPISP	NN	Prediction based on the solvent accessibility and PSSM of spatially neighboring surface residues	https://pipe.rcc.fsu.edu/ppisp.html	Structure-based	141
CPORT	Scoring function	Combine multiple individual predictors (WHISCY, PIER, ProMate, cons-PPISP, SPPIDER, and PINUP) to generate a consensus prediction. Specifically aimed at the use of interface predictions in data-driven docking with HADDOCK	https://milou.science.uu.nl/services/CPORT	Structure-based meta-server	142
DPPI	Siamese-like convolutional NN	Model prediction based on high-quality experimental PPIs data and evolutionary information of protein pairs	https://github.com/hashemifar/DPPI	Sequence-based	143
ECLAIR (Interactome INSIDER)	Ensemble of RFs	A model trained on a different set of features including biophysical, structure-based, docking-based and co-evolution features	http://interactomeinsider.yulab.org	Structure and co-evolution based	144

Evcomplex	Maximum entropy	Prediction of inter-residue contacts from multiple sequence alignments	https://evcouplings.org/complex	Co-evolution based	145
InterProSurf	Scoring functions	The method is based on SASA of residues in the isolated subunits, propensity of interface residues and residues in surface regions with high interface propensities	http://curie.utmb.edu/prosurf.html	Structure-based	146
ISPRED4	SVM with grammar-based correction	The model has been trained using features extracted from the protein sequence and structure	https://ispred4.biocomp.unibo.it/ispred	Structure- and sequence-based	137
meta-PPISP	Linear regression	Build based on three independent servers' cons-PPISP, PINUP, and Promate	http://pipe.scs.fsu.edu/meta-ppisp.html	Structure-based meta-server	147
PAIRpred	Multiple pairwise kernel SVM	Uses structure (SASA, residue depth, half-sphere amino acid composition and a protrusion index), and sequence (profile features from the PSSMs and predicted RASA) to predict protein-protein interactions	http://combi.cs.colostate.edu/supplements/pairpred	Sequence- and structure-based	148
PIER	PLS regression	The model uses solvent accessibility and evolutionary conservation to predict interfaces from a single protein structure	http://abagyan.ucsd.edu/PIER	Structure-based	149
PINUP	Empirical scoring function	Effective residue-energy score, accessible-surface-area dependent interface-propensity, and residue conservation score are used to train a model for binding site prediction of monomeric proteins	N/A	Structure-based	150
PPiPP	An ensemble of 24 NNs	It uses a binary encoding of 20 amino acids and PSSM to predict pairwise binding sites	N/A	Sequence-based	151

Continued

Table 2 Alphabetically ordered list of ML-based methods and web-servers for prediction and identification of PPIs.—cont'd

Methods name	Type of ML algorithm	Main features	Server or meta server URL	Type of model	Reference
PPI_SVM	Two-class SVM	Physical interactions of constituent domains of protein pairs extracted from the Database of Interacting Proteins	N/A	Structure-based	152
PredPPIS	SVM and Bayesian classifiers	Combine 36 sequence features divided into 5 categories (orthogonal amino acid indices, PSSM profiles, predicted secondary structures, tendency of being located on disordered regions and sequence conservation) for each amino acid residue	http://bsaltools.ym.edu.tw/predppis	Sequence-based	153
PresCont	SVM	SASA, hydrophobicity, conservation and the local environment of each amino acid on the protein surface deduced from the 3D structure of an individual protein and an MSA composed of homologous	https://bioinf.ur.de/prescont.php	Structure-based	154
PredUS	SVM	The model uses a structural alignment method to identify structural neighbors of a given protein. The interface of the structural neighbor is then mapped against the query protein to predict if a surface residue is at an interface	https://bhapp.c2b2.columbia.edu/PredUs	Structural homology-based	155
PRISM	Scoring function	PRISM predicts binding residues by using geometric complementarity and evolutionary conservation of hot spots	http://cosbi.ku.edu.tr/prism	Structure-based	156
PROFisis (earlier ISIS)	NN	Evolutionary profiles along with predictions of solvent accessibility and secondary structure were used to predict whether a residue is likely to be part of a protein-protein interface	https://www.predictprotein.org	Sequence-based	157

ProMate	Composite probability	Defined patch encompassed by circles with a radius of 10 Å drawn from a center of a given protein surface residue is obtained. Using features that differentiate binding from non-binding surfaces, the model evaluates the probability of each circle to appear at the interface	http://bioinfo41.weizmann.ac.il/promate/promate.html	Structure-based	158
PSIVER	Naïve Bayes classifier	The model predicts PPI sites using a PSSM and predicted accessibility	https://mizuguchilab.org/PSIVER	Sequence-based	135
SeRenDIP	RF	Sequence features such as conservation, secondary structure, ASA/RSA, protein size, backbone flexibility and sequence specificity were used for protein interface prediction	http://www.ibi.vu.nl/programs/serendipwww	Sequence-based	138
SHARP2	Scoring function	Solvation potential, hydrophobicity, ASA, residue interface propensity, planarity and protrusion features were included in the model	N/A	Structure-based	159
SPPIDER	Combined output of 10 NNs	Integrates relative solvent accessibility (RSA) with high-detailed structural data as features to predict PPI sites	http://sppider.cchmc.org	Structural- and sequence-based	136
UNISPPI	Decision tree	Use of 20 combinations of amino acids frequencies from interacting and non-interacting proteins	N/A	Sequence-based	160
WHISCY	Scoring function	A model that combines surface conservation and structural information to predict protein-protein interfaces	https://milou.science.uu.nl/services/WHISCY	Structure-based	161

Abbreviations: DL, deep learning; ML, machine learning; MSA, multiple sequence alignment; N/A, not available; NN, neural network; PLS, partial least square; PPI, protein-protein interaction; PSSM, position-specific scoring matrix; RASA, relative accessible surface area; RF, random forest; SASA, solvent-accessible surface area; SVM, support vector machine; XGBoost, extreme gradient boosting.

Table 3 Alphabetically ordered list of available public databases and meta databases of protein-protein interactions.

Method name	Main features	Website URL	Total interactions (as of 22nd October)[a]	Reference
APID	A comprehensive and curated collection of protein interactomes for >400 organisms; also includes 90,379 distinct proteins. Integrates data from databases of interactions (BIND, BioGRID, DIP, HPRD, IntAct, MINT) and also from 3D structures	http://apid.dep.usal.es	678,441	162
BioGRID	Database that includes curated information from physic and genetic interactions	https://orcs.thebiogrid.org	1,598,688	163
CPDB	Integrates interaction networks in *Homo sapiens*; including binary and complex protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target interactions. Data is originated from 32 public resources for interactions and curated from the literature	http://cpdb.molgen.mpg.de	660,318	164
DIP	Database catalogs of experimentally determined interactions between proteins, both manually and computational curated	https://dip.doe-mbi.ucla.edu/dip/	81,923	165
GPCRdb	Database centralizing many of the known structural information on GPCRs, also makes available a large set of tools for GPCR handling and analysis	https://www.gpcrdb.org/	N/A	23,166
gpDB	Database of GPCRs, G-proteins, effectors and their interactions	http://bioinformatics.biol.uoa.gr/gpDB/	N/A	167
HPRD	Human Protein Reference Database that includes manually curated information from post-translational modifications, interaction networks and disease association for each protein	http://www.hprd.org/	41,327	168

IID	Database that integrates tissue-specific PPIs for model organisms and human	http://iid.ophid.utoronto.ca	4,927,742	169
InnateDB	Contain information about interactions involved in mammalian innate immunity. Integration of curated interactions from several databases (IntAct, DIP, MINT, BIND and BioGRID) and the literature	https://www.innatedb.com/	829,948	170
IntAct	Database and analysis tools for molecular interaction data. Contain manually curated information from the literature or by direct submission	https://www.ebi.ac.uk/intact/	960,621	171
Interactome3D	Structural annotation of PPIs networks. Visualization and download of structural information from protein interactions. Contains information from other PPIs databases (3did, BIND, BioGRID, DIP, HPRD, InnateDB, IntAct)	https://interactome3d.irbbarcelona.org	239,859	172
iRefWeb	Database of protein-protein interactions (PPI) consolidated from major public databases (BIND, BioGRID, CORUM, DIP, IntAct, HPRD, MINT, MPact, MPPI, OPHID)	http://wodaklab.org/iRefWeb/	N/A	173
Mentha	Database of manually curated molecular interactions from diverse databases (BioGRID, DIP, IntAct, MatrixDB, MINT)	https://mentha.uniroma2.it	741,337	174
MINT	Database of experimentally verified protein-protein interactions mined from the literature	https://mint.bio.uniroma2.it	68,501	175
MIPS	Mammalian PPIs curated database from the scientific literature	http://mips.helmholtz-muenchen.de/proj/ppi/	N/A	176

Continued

Table 3 Alphabetically ordered list of available public databases and meta databases of protein-protein interactions.—cont'd

Method name	Main features	Website URL	Total interactions (as of 22nd October)[a]	Reference
Negatome	Database of experimentally supported non-interacting protein pairs, derived from manual curation of literature and by the analyses of protein complexes from the PDB	http://mips.helmholtz-muenchen.de/proj/ppi/negatome/	N/A	177
PICKLE	Protein InteraCtion KnowLedgebasE is a metadatabase that integrates publicly available PPIs via genetic information ontology. PICKLE combines information from BioGRID, IntAct, HPRD, MINT and DIP	http://www.pickle.gr	191,510	178
PINA	Protein Interaction Network Analysis allows the construction of PPI networks through the inclusion of data from six public PPI databases (IntAct, MINT, BioGRID, DIP, HPRD and MIPS MPact)	http://omics.bjcancer.org/pina/	N/A	179
PrePPI	Database of predicted (Bayesian framework that combines structural, functional, evolutionary and expression information) and experimentally determined PPIs for the human proteome	https://bhapp.c2b2.columbia.edu/PrePPI/	~1,350,000	180
STRING	Database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations	https://string-db.org/	~3,123,056,667	181

[a]N/A, not applicable.

Nevertheless, when developing such models, it is essential to consider that most of the time, in MP interface prediction tasks, we have an imbalanced class distribution, i.e., the majority class is represented by non-interface residues while interface residues are the less representative class, thus influencing model predictive power. In addition to interfacial residues features, several reviews[132] and methods such as PPiPP,[151] PS-HomPPI,[183] PAIRpred,[148] or BIPSPI[139] have proven the importance of including partner-specific information for interface prediction (Table 2). To date, the most commonly used interfacial residues features are divided into three broad categories depending on the type of the included information: structure-, sequence-, and co-evolutionary-based.

(i) *Structure-based features*: the most accurate and used information about residues at PPI interfaces arose from the 3D structure of protein complexes (experimentally obtained or by using homology modeling methods). These features also include, but are not limited to, solvent-accessible surface area (SASA),[127] crystallographic B-factors,[184] or secondary structure.[185] An essential characteristic of interfaces is the presence of hot spots; cooperative and highly conserved residues with a significant degree of chemical and spatial complementarity that have a major contribution to the binding affinity.[186] To date, not only some ML structure but also sequence-based models have been developed to predict hot spots such as the ensemble SpotON[123] or the SVM-based KFC2,[187] PREDHS2,[188] PSIPRED,[189] SBHD[190] and POCKETQUERY.[191] Some representative structure-based ML methods are presented in Table 2. One of the disadvantages of structure-based features is that they are dependent on the available solved 3D structures, which is a current problem (as previously mentioned), particularly in the case of MPs. Secondly, most of the structural information of interface residues are retrieved from proteins in the unbound state (*apo* form), which can be different from the one obtained after protein-protein complexes formation.[132] Hence, the use of other features, such as the one based on protein sequences, is a viable option to counteract this problem.

(ii) *Sequence-based features*: characteristics that are derived from the amino acid sequence of the protein. Examples of these type of features can include the ones extracted directly from the sequence, such as amino acid composition and the corresponding neighboring residues, propensity values,[192] physicochemical properties (e.g., hydrophobicity, polarity, charge)[193] or the predicted ones, like the relative accessible surface area (RASA).[194] Additionally, some methods that take as input

sequence profiles were also explicitly developed for the characterization of MP, such as the LIPid-facing Surface (LIPS)[195] for prediction of helix-lipid interfaces of TM helices or MEMPACK[196] and TMhit,[197] both using SVMs, TMHcon[198] using ANNs, and TMhhcp[199] using RFs for prediction of MPs helix-helix contacts among other features. MemBrain,[194] an ML-based method, was also recently developed for transmembrane helices, residue-residue contacts, and RASA prediction with an accuracy of 97.9%. Representative sequence-based ML methods are presented in Table 2. The usage of sliding windows, which consists of analyzing a set of residues sequentially related to the target amino acid, has been used as a way to emulate the structural dependency of primary structure proximal residues. However, this fails to acknowledge distant sequence amino acids than can, nevertheless, have three-dimensional close distances. Hence, this method can improve performance, although most of the sequence-based models have lower performance than structure-based models. Despite that, these features provide valuable information for oligomers interface prediction and in most cases, are used together with other features such as structure- or evolutionary-based.

(iii) *Co-evolutionary based features*: although it is a feature derived from the protein sequence, due to the complexity of methods involved in the study of evolutionary scores, we will consider them as a separated group. It was demonstrated that oligomer interfacial residues are typically more conserved among homologous proteins and more likely to co-evolve when compared with non-interfacial surface residues, mainly to preserve the interaction interface and consequently protein function.[101] The degree of conservation can be assessed by developing position-specific scoring matrices (PSSMs) from MSAs comparing each amino acid of the query sequence to the corresponding sequence of its homologous (Fig. 2). High-quality MSAs can provide useful information concerning correlated mutations that can be used for training ML algorithms (Table 2)[126,203,204] (for more information about the importance of sequence evolutionary data, please see Nicoludis and Gaudet review).[205] Another interesting approach is the assessment of residue pairs and their joint conservation analysis, which culminates in full 3D structure prediction from the protein sequence, and displays relevant results, considering that only co-evolutionary features were used.[206] Despite the success of co-evolutionary analysis, the presence of multiple paralogs protein families (out of 2985 Pfam31 families,[207] 2244 have a

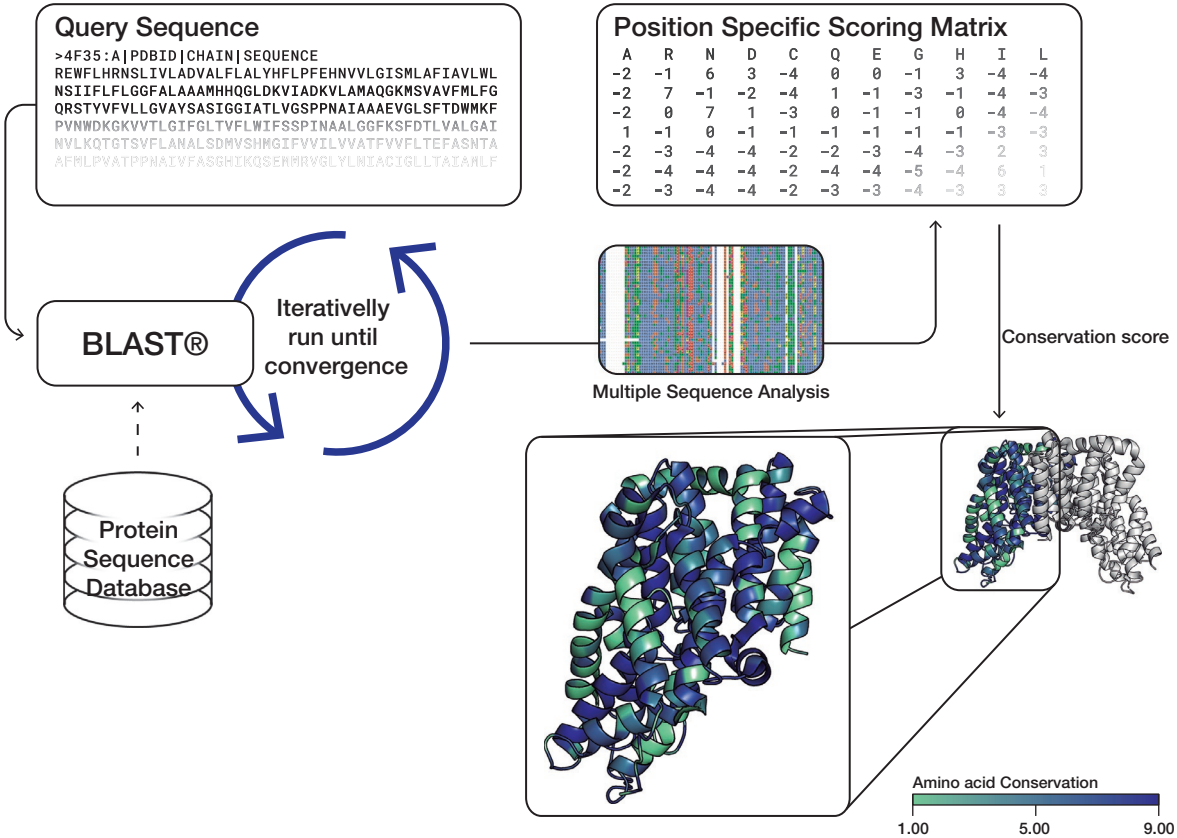

Fig. 2 Pipeline for the assessment of protein amino acid conservation scores. A multiple sequence alignment of a query protein and a homologous non-redundant protein database is used to produce a Position Specific Scoring Matrix (PSSM) using PSI-BLAST.[200] The conservation score obtained as output from PSSM corresponds to each amino acid (or gap) at each position of the alignment. In the right-hand square below is the bacterial dicarboxylate/sodium symporter (PDB id: 4F35)[201] colored by amino acid conservation using ConSurf.[202]

mean of more than three paralogs per species, and 1093 a mean above 5) in MSA alignments could be a challenge for understanding PPIs, particularly in the case of eukaryotes.[208–210] More recently, Wu and colleagues have suggested the use of metagenome sequence data as a complement to MSA for residue contact prediction in the case of proteins families with an insufficient number of homologs.[211]

Contrarily to non-MPs, in the last years, very few ML methods were explicitly developed for MPs oligomers interface prediction. To date, only RFs,[212,213] SVMs[214] and NNs[215] approaches have been applied for this purpose. Bordner was one of the pioneer studies that used a RF classifier to predict the interface residues of transmembrane proteins, although the final dataset included more α-helical than β-barrel structures; a problem which can be traced back to the fact that this impairment is also noticeable at experimentally determined structures level.[212] The method was trained using sequence-based features of individual intramembrane surface residues (including only the ones with a relative SASA $\geq$0.2), such as their overall distribution and evolutionary conservation (MSA), reaching a prediction performance of area under the curve (AUC) of 0.75. This study also highlights the importance of developing separated predictors for membrane and non-membrane proteins to achieve better model performance.[212] More recently, by using separate RF models, Zeng and colleagues[213] have developed the MBPred (Membrane-protein Binding-residues Prediction), a method that predicts interface residues in transmembrane (MBPredTM), cytoplasmic (MBPredCyto) and extracellular segments (MBPredExtra) as well as in the entire amino acid sequence (MBPredAll). Contrarily to Bordner,[212] this work only used a dataset of non-redundant α-helical membrane proteins, although they included more features either from the primary sequence (relative position, physical properties and segment) or from MSAs (residue conservation, evolutionary profile, PSSM), cumulative and maximum co-evolution strength, and lipid accessibility). MBPred achieved a slightly better AUC than the previous model and other ones trained on globular proteins (0.79 and 0.73 on the cross-validation and independent test dataset, respectively).[213] Using a different approach, Li and colleagues[215] developed a neural network-based method able to predict interface residues of a non-redundant set of oligomers of α-helical integral MPs, with an AUC of 0.75. In contrast with previous studies, one of the advantages of this method is that it calculates the weighted contact numbers—WCNs (number of its neighboring residues weighted by spatial proximity)[216] of surface residues and use these values for the prediction of interfaces.[215]

The construction of ML-based predictors has to face a set of challenges, some of which have been previously stated. Many of these issues are deepened when considering MPs in particular, and some stem from the already difficult task of experimental determination. Some of the current challenges that must be considered when developing future ML-based models for MPs interface prediction are:

(i) Data availability:
- **a.** restriction in the number of available experimentally determined structures;
- **b.** deposited structures lack of variety: α-helix vs β-barrel impairment;
- **c.** structures' reliability: low resolution and poor membrane insertion;
- **d.** incomplete structures: α-carbon only structures, protein sequences with gaps and uncertain regions;
- **e.** non-biologically accurate structures: common lack of hydrogens.

(ii) Class definition:
- **a.** determined on the same structures affected by the issues listed in point (i);
- **b.** in some cases, non-consensual, leading to non-comparable results.

(iii) Features:
- **a.** there is no sure way to know beforehand what is the individual feature contribution to model performance, hence, the features must be extracted under the same conditions and the models must be tested, which is computationally expensive;
- **b.** some features may display erratic behavior by having no connection to the class, but being apparently representative of it;
- **c.** the concatenation of features must include the biological knowledge on the subject, otherwise, this can lead to deceptively good results, by, for example, including features that are directly related to the class.

Several more problems than the stated above, arise when dealing with ML-based MP interface prediction. On the bright side, some of the previously mentioned difficulties also stand, as the answer as to why computational methods are a viable approach to deal with challenges that, on a single sample point of view, are seemingly unsolvable. With the rise of ML (Deep Learning (DL) in particular), the handling of big data has evolved from being a hindrance to a remarkable advantage. By evaluating multiple structures on the same algorithm, it is possible to dilute the issues of particular structures and residues that arise from experimental methodology into all-encompassing algorithms that, when appropriately programmed, can automatically distinguish between useful and non-reliable information.

The usage of ANN, in particular, opens the possibility of embedded feature extraction, since the non-contributing features can be iteratively filtered out. Seemingly, this process is also able to set uniform standards for all features, minimizing the issues that arise from features extracted or engineered from different sources and by different authors. Thus, nowadays, using the computational power that less than a decade ago would be insufficient, it is possible to build very robust algorithms to which one of the main factors is the amount of available data.

3. Targetting PPIs: Orthosteric and allosteric modulation

Targeting PPIs has become a promising strategy in drug discovery since they display a key role in both several biological processes and pathological conditions. Nevertheless, the large and flat interfaces of PPIs make this achievement a challenging task, mostly due to the lack of drug-binding pockets.[217,218]

Two main approaches can be applied for targeting PPIs: orthosteric and allosteric modulation, both by using small-molecules or peptidomimetic agents.[218,219] While orthosteric PPI modulators hinder or stabilize the interaction between both protein partners by binding to the PPI interface, allosteric agents bind to a different location from the native binding site, triggering conformational changes that are ultimately responsible for also preventing or stabilizing PPIs.[219]

3.1 Orthosteric modulation

To date, orthosteric modulation by small-molecules is the primary strategy pursued to target PPIs.[220] Orthosteric modulators generally display different physicochemical properties (e.g., a higher molecular size) in comparison with inhibitors of enzymes or receptors, mostly because of the nature and shape of protein-protein interfaces.[221] These type of ligands are particularly suitable for targeting small PPIs and/or those in which hot-spot residues define proper binding sites, as the absence of concavity in protein-protein interfaces narrows the ligand contacts with PPI surface, limiting its tight binding.[220]

Since the interaction of a small-molecule with hot-spot residues can compete with the binding protein partner, disrupting PPI by targeting hot spots has been a widely applied strategy able to identify potent and selective PPI inhibitors.[222–225] Therefore, the identification of hot-spot residues

works as a way to identify suitable inhibitor binding sites at protein-protein interfaces.[226] MDs simulations and docking studies, as well as ML methods (reviewed in the previous section), are commonly applied in hot-spot identification.[43] Once identified, hot-spot knowledge can be combined with other computational techniques to identify druggable binding sites at protein-protein interfaces, or even PPI modulators. An example of a successful multistep approach that combines hot-spot prediction, pharmacophore screening, and molecular docking was the identification of IFNA-IFN receptor interaction inhibitors. By using iPred, which is a knowledge-based scoring function tool for hot-spot prediction, and a structure-based pharmacophore approach, a new promising inhibitor targeting the IFNA-IFN interface was identified.[43]

A new approach, based on the development of covalent inhibition, has emerged as a more efficient promising strategy for targeting PPIs.[217] This approach involves the covalent modification of a nucleophilic residue, such as cysteine or methionine, located at nearest-neighbor PPI interfaces. The main objective is to achieve an efficient and prolonged target modulation, ultimately able to overcome the drawback of large and featureless PPI interfaces. Examples of such successful modulators are the inhibition of KRAS G12C, MCL-1, and BRD4 proteins through covalent modification of cysteine, lysine, and methionine, respectively.[217]

Among the small-molecule PPI inhibitors identified, some have entered clinical trials (including some covalent inhibitors), and few of them were approved.[226] Gabapentin, which is a GABA mimetic drug, firstly used in the treatment of epilepsy, was reported as a competitive inhibitor of the interaction between the GPCR bradykinin and prokineticin 2, reducing protein kinase C epsilon (PKCε) translocation and ultimately leading to a relief of neuropathic pain.[227]

It is important to note that some antibodies and peptides were also identified as PPI inhibitors since they display larger interacting interfaces.[226] A peptide inhibitor of LMP-1 oligomerization was recently identified, representing a starting point for EBV targeting by inhibition of LMP-1's TMD5 trimerization. As far as we know, this molecule represents the first peptide inhibitor involved in the disruption of homotrimeric transmembrane helices.[228] Nevertheless, antibodies and peptides usually display lower bioavailability in comparison with small-molecules, which often make them unattractive candidates for PPIs modulation.

Nanobodies (Nbs) have emerged as a new class of promising antibody-based therapeutics able to overcome the limitations of antibodies. Nbs usually

displaying high selectivity and extended half-lives when compared with small-molecules.[229] These molecules were reported as targeting GPCRs, namely chemokine receptors. Nbs 238D2 and 238D4 are both able to bind to CXCR4 protein, competing for the binding with CXCL12. This approach prevented CXCL12-dependent binding and signaling, inhibiting HIV-1 replication *in vitro*.[230]

3.2 Allosteric modulation

Allosteric modulation is mainly applied when the PPI interface is large and flat, being these types of modulators more drug-like than orthosteric PPIs ones.[218] Moreover, allosteric ligands can overcome one of the most significant issues of orthosteric binding, the competition with the bulky PPI partners, which usually display a higher affinity for the protein-protein interface.[219] Several allosteric PPIs modulators have been identified so far. Among them, synthetic peptides developed taking into account the structure of TM helices from bovine rhodopsin (Rho), have been used to study the Rho dimer interface, including the effect of its disruption.[231] Another example of allosteric modulation involves NTS1R-D2R complexes. It was suggested that the binding of bivalent ligand CS148 to both NTS1R and D_2R protomers (as agonist and antagonist, respectively) promotes a conformational change in NTS1R-D_2R interface, which results in a calcium response comparable to NTS1R monomer activation.[232]

Altogether, and despite the identification of PPIs modulators by different experimental and computational tools for a large number of diverse complexes, further studies are needed to identify new and more selective PPIs inhibitors involved in oligomerization. Information about PPI modulators can be found in several databases, such as TIMBAL,[233] 2P2I,[234] and iPPI-DB,[235,236] which compiles information about the nature and structure of those molecules.

4. Concluding remarks

GPCRs are a broad family of membrane receptors that have an essential role in multiple diseases, and because of that are targets of about 34% of total approved drugs. Diverse experimental and computational approaches have demonstrated the existence of GPCR dimers and high-order oligomers and their impact on GPCR function and dynamics. The discovery of GPCR dimers with a physiological importance suggests that new and more targeted drugs can be developed by targeting these structures.

In this work, we review experimental, computational, and ML methods that are used to characterize and predict oligomerization interaction interfaces in GPCRs. Among the experimental methods, fluorescence-based assays are the most used to study PPIs, being often applied in combination with affinity-based methods and/or genetic assays. More recent techniques, such as TIRFM and single-molecule imaging, gave further insights into the oligomerization state of GPCRs, while a new FRET-based strategy focused on ligand binding selectivity for hetero-oligomers was also identified. Concerning computational methods, structure-based methods are preferred over sequence-based ones when 3D information about the dimer/oligomer is available. Additionally, CGMD is usually the first choice to study the GPCR dynamics engaging PPIs, due to the large size of the oligomeric systems and the long length scales needed to observe reliable conformational changes.

More recently, diverse structure- and/or sequence-based models that use ML algorithms appear to support the prediction of PPIs interface. Although these models can also be used for GPCR interface dimers prediction, the complex interaction and modulation between these receptors and the hydrophobic lipid membrane is hindering this application. Some methods that combine structure-, sequence- and co-evolution-based features with ML algorithms were explicitly developed for membrane proteins; however, the development of a suitable method for GPCR dimers interface prediction is still needed.

Regarding targeting of PPIs interface, both orthosteric and allosteric modulators have been identified. Although new promising PPIs modulators have emerged (e.g., covalent inhibitors, Nbs), further studies are needed to identify more selective and safer PPI modulators.

Although different strategies are applied in the prediction and targeting of PPIs, all methods identified, so far, present limitations. Therefore, the combination of experimental and computational and/or ML methods have been applied to overcome the drawbacks of each technique, leading to a better prediction and characterization of GPCR PPIs.

Nevertheless, much more work needs to be done to have a holistic understanding of GPCR oligomerization and better targeting the interfaces in GPCR oligomers.

Acknowledgments

I.S.M. was funded by the Fundação para a Ciência e a Tecnologia (FCT) Investigator programme—IF/00578/2014 (co-financed by European Social Fund and Programa Operacional Potencial Humano). This research was funded by the European Regional

Development Fund (ERDF), through the Centro 2020 Regional Operational Programme under project CENTRO-01-0145-FEDER-000008: BrainHealth 2020, and through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation and Portuguese national funds via FCT, under projects POCI-01-0145-FEDER-007440 and PTDC/QUI-OUT/32243/2017. C.A.V.B. was supported by FCT through PhD scholarship SFRH/BD/145457/2019.

References

1. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov*. 2017;16(12):829–842. https://doi.org/10.1038/nrd.2017.178.
2. Moreira IS. Structural features of the G-protein/GPCR interactions. *Biochim Biophys Acta Gen Subj*. 2014;1840(1):16–33. https://doi.org/10.1016/j.bbagen.2013.08.027.
3. Lemos A, Melo R, Preto AJ, Almeida JG, Moreira IS, Dias Soeiro Cordeiro MN. In silico studies targeting G-protein coupled receptors for drug research against Parkinson's disease. *Curr Neuropharmacol*. 2018;16(6):786–848. https://doi.org/10.2174/1570159x16666180308161642.
4. Kniazeff J, Prézeau L, Rondard P, Pin JP, Goudet C. Dimers and beyond: the functional puzzles of class C GPCRs. *Pharmacol Ther*. 2011;130(1):9–25. https://doi.org/10.1016/j.pharmthera.2011.01.006.
5. Milligan G. G protein-coupled receptor hetero-dimerization: contribution to pharmacology and function. *Br J Pharmacol*. 2009;158(1):5–14. https://doi.org/10.1111/j.1476-5381.2009.00169.x.
6. Bouvier M, Hébert TE. CrossTalk proposal: weighing the evidence for Class A GPCR dimers, the evidence favours dimers. *J Physiol*. 2014;592(12):2439–2441. https://doi.org/10.1113/jphysiol.2014.272252.
7. Salahpour A, Angers S, Mercier JF, Lagacé M, Marullo S, Bouvier M. Homodimerization of the β2-adrenergic receptor as a prerequisite for cell surface targeting. *J Biol Chem*. 2004;279(32):33390–33397. https://doi.org/10.1074/jbc.M403363200.
8. Dupré DJ, Robitaille M, Éthier N, Villeneuve LR, Mamarbachi AM, Hébert TE. Seven transmembrane receptor core signaling complexes are assembled prior to plasma membrane trafficking. *J Biol Chem*. 2006;281(45):34561–34573. https://doi.org/10.1074/jbc.M605012200.
9. Sleno R, Hébert TE. The dynamics of GPCR oligomerization and their functional consequences. In: Shukla AK, ed. *International Review of Cell and Molecular Biology*. Elsevier Inc.; 2018:141–171. G Protein-Coupled Receptors: Emerging Paradigms in Activation, Signaling and Regulation Part A. vol. 338. https://doi.org/10.1016/bs.ircmb.2018.02.005
10. Dupré DJ, Hébert TE. Biosynthesis and trafficking of seven transmembrane receptor signalling complexes. *Cell Signal*. 2006;18(10):1549–1559. https://doi.org/10.1016/j.cellsig.2006.03.009.
11. Dupré DJ, Robitaille M, Rebois RV, Hébert TE. The role of Gβγ subunits in the organization, assembly, and function of GPCR signaling complexes. *Annu Rev Pharmacol Toxicol*. 2009;49(1):31–56. https://doi.org/10.1146/annurev-pharmtox-061008-103038.
12. Gavalas A, Lan TH, Liu Q, Corrêa IR, Javitch JA, Lambert NA. Segregation of family A G protein-coupled receptor protomers in the plasma membrane. *Mol Pharmacol*. 2013;84(3):346–352. https://doi.org/10.1124/mol.113.086868.
13. Kawano K, Yano Y, Omae K, Matsuzaki S, Matsuzaki K. Stoichiometric analysis of oligomerization of membrane proteins on living cells using coiled-coil labeling and spectral imaging. *Anal Chem*. 2013;85(6):3454–3461. https://doi.org/10.1021/ac400177a.

14. Felce JH, Knox RG, Davis SJ. Type-3 BRET, an improved competition-based bioluminescence resonance energy transfer assay. *Biophys J*. 2014;106(12):L41–L43. https://doi.org/10.1016/j.bpj.2014.04.061.
15. Scarselli M, Annibale P, McCormick PJ, et al. Revealing G-protein-coupled receptor oligomerization at the single-molecule level through a nanoscopic lens: methods, dynamics and biological function. *FEBS J*. 2016;283(7):1197–1217. https://doi.org/10.1111/febs.13577.
16. Guidolin D, Marcoli M, Tortorella C, Maura G, Agnati LF. Receptor-receptor interactions as a widespread phenomenon: novel targets for drug development? *Front Endocrinol (Lausanne)*. 2019;10:53. https://doi.org/10.3389/fendo.2019.00053.
17. White SH. Biophysical dissection of membrane proteins. *Nature*. 2009;459(7245):344–346. https://doi.org/10.1038/nature08142.
18. Saier MH. TCDB: the transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res*. 2006;34(90001):D181–D186. https://doi.org/10.1093/nar/gkj001.
19. Kozma D, Simon I, Tusnády GE. PDBTM: protein data bank of transmembrane proteins after 8 years. *Nucleic Acids Res*. 2013;41(D1):D524–D529. https://doi.org/10.1093/nar/gks1169.
20. Tusnády GE, Dosztányi Z, Simon I. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics*. 2005;21(7):1276–1277. https://doi.org/10.1093/bioinformatics/bti121.
21. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res*. 2012;40(D1):D370–D376. https://doi.org/10.1093/nar/gkr703.
22. Newport TD, Sansom MSP, Stansfeld PJ. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res*. 2019;47(D1):D390–D397. https://doi.org/10.1093/nar/gky1047.
23. Pándy-Szekeres G, Munk C, Tsonkov TM, et al. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res*. 2018;46(D1):D440–D446. https://doi.org/10.1093/nar/gkx1109.
24. Khelashvili G, Dorff K, Shan J, et al. GPCR-OKB: the G protein coupled receptor oligomer knowledge base. *Bioinformatics*. 2010;26(14):1804–1805. https://doi.org/10.1093/bioinformatics/btq264.
25. Zhang J, Yang J, Jang R, Zhang Y. GPCR-I-TASSER: a hybrid approach to G protein-coupled receptor structure modeling and the application to the human genome. *Structure*. 2015;23(8):1538–1549. https://doi.org/10.1016/j.str.2015.06.007.
26. Borroto-Escuela DO, Brito I, Romero-Fernandez W, et al. The G protein-coupled receptor heterodimer network (GPCR-HetNet) and its hub components. *Int J Mol Sci*. 2014;15(5):8570–8590. https://doi.org/10.3390/ijms15058570.
27. Townsend-Nicholson A, Altwaijry N, Potterton A, Morao I, Heifetz A. Computational prediction of GPCR oligomerization. *Curr Opin Struct Biol*. 2019;55:178–184. https://doi.org/10.1016/j.sbi.2019.04.005.
28. Warne T, Serrano-Vega MJ, Baker JG, et al. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature*. 2008;454(7203):486–491. https://doi.org/10.1038/nature07101.
29. 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(4):419–425. https://doi.org/10.1038/nsmb.2504.
30. Wu B, Chien EYT, Mol CD, et al. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*. 2010;330(6007):1066–1071. https://doi.org/10.1126/science.1194396.

31. Thompson AA, Liu W, Chun E, et al. Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature*. 2012;485(7398):395–399. https://doi.org/10.1038/nature11085.
32. Morimoto K, Suno R, Hotta Y, et al. Crystal structure of the endogenous agonist-bound prostanoid receptor EP3. *Nat Chem Biol*. 2019;15(1):8–10. https://doi.org/10.1038/s41589-018-0171-8.
33. Robertson N, Rappas M, Doré AS, et al. Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature*. 2018;553(7686):111–114. https://doi.org/10.1038/nature25025.
34. Cao C, Tan Q, Xu C, et al. Structural basis for signal recognition and transduction by platelet-activating-factor receptor. *Nat Struct Mol Biol*. 2018;25(6):488–495. https://doi.org/10.1038/s41594-018-0068-y.
35. Park JH, Scheerer P, Hofmann KP, Choe H-W, Ernst OP. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature*. 2008;454(7201):183–187. https://doi.org/10.1038/nature07063.
36. Nakamichi H, Buss V, Okada T. Photoisomerization mechanism of rhodopsin and 9-cis-rhodopsin revealed by x-ray crystallography. *Biophys J*. 2007;92(12): L106–L108. https://doi.org/10.1529/biophysj.107.108225.
37. Standfuss J, Xie G, Edwards PC, et al. Crystal structure of a thermally stable rhodopsin mutant. *J Mol Biol*. 2007;372(5):1179–1188. https://doi.org/10.1016/j.jmb.2007.03.007.
38. Wang C, Wu H, Katritch V, et al. Structure of the human smoothened receptor bound to an antitumour agent. *Nature*. 2013;497(7449):338–343. https://doi.org/10.1038/nature12167.
39. Koehl A, Hu H, Feng D, et al. Structural insights into the activation of metabotropic glutamate receptors. *Nature*. 2019;566(7742):79–84. https://doi.org/10.1038/s41586-019-0881-4.
40. Glukhova A, Thal DM, Nguyen AT, et al. Structure of the adenosine A1 receptor reveals the basis for subtype selectivity. *Cell*. 2017;168(5):867–877.e13. https://doi.org/10.1016/j.cell.2017.01.042.
41. Wu H, Wacker D, Mileni M, et al. Structure of the human κ-opioid receptor in complex with JDTic. *Nature*. 2012;485(7398):327–332. https://doi.org/10.1038/nature10939.
42. Muto T, Tsuchiya D, Morikawa K, Jingami H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. *Proc Natl Acad Sci USA*. 2007;104(10):3759–3764. https://doi.org/10.1073/pnas.0611577104.
43. Schiedel AC, Kose M, Barreto C, et al. Prediction and targeting of interaction interfaces in G-protein coupled receptor oligomers. *Curr Top Med Chem*. 2018;18(8):714–746. https://doi.org/10.2174/1568026618666180604082610.
44. Fernandez-Duenas V, Llorente J, Gandia J, et al. Fluorescence resonance energy transfer-based technologies in the study of protein-protein interactions at the cell surface. *Methods*. 2012;57(4):467–472. https://doi.org/10.1016/j.ymeth.2012.05.007.
45. Guo H, An S, Ward R, et al. Methods used to study the oligomeric structure of G-protein-coupled receptors. *Biosci Rep*. 2017;37(2):1–20. https://doi.org/10.1042/BSR20160547. Portland Press Ltd.
46. Fernandez-Duenas V, Gomez-Soler M, Valle-Leon M, Watanabe M, Ferrer I, Ciruela F. Revealing adenosine A2A-dopamine D2 receptor heteromers in Parkinson's disease post-mortem brain through a new alphascreen-based assay. *Int J Mol Sci*. 2019;20(14):3600. https://doi.org/10.3390/ijms20143600.
47. Dagher OK, Jaffa MA, Habib A, Ziyadeh FN, Jaffa AA. Heteromerization fingerprints between bradykinin B2 and thromboxane TP receptors in native cells. *PLoS One*. 2019;14(5): e0216908, https://doi.org/10.1371/journal.pone.0216908.

48. Heuninck J, Hounsou C, Dupuis E, et al. Time-resolved FRET-based assays to characterize G protein-coupled receptor hetero-oligomer pharmacology. In: Tiberi M, ed. *G Protein-Coupled Receptor Signaling*. New York, NY: Humana Press; 2019: Methods in Molecular Biology. vol. 1947.
49. Jiang Y, Yuan Y, Zhang X, et al. Use of network model to explore dynamic and allosteric properties of three GPCR homodimers. *RSC Adv*. 2016;6(108):106327–106339. https://doi.org/10.1039/c6ra18243g.
50. Meng X-Y, Mezei M, Cui M. Computational approaches for modeling GPCR dimerization. *Curr Pharm Biotechnol*. 2014;15(10):996–1006. https://doi.org/10.2174/1389201015666141013102515.
51. Kaczor AA, Bartuzi D, Stępniewski TM, Matosiuk D, Selent J. Protein–protein docking in drug design and discovery. In: Gore M, Jagtap U, eds. *Computational Drug Discovery and Design*. New York, NY: Humana Press; 2018: Methods in Molecular Biology. vol. 1762.
52. Kaczor AA, Selent J, Poso A. Structure-based molecular modeling approaches to GPCR oligomerization. In: Michael Conn P, ed. *Methods in Cell Biology*. vol. 117. Academic Press; 2013.
53. Shukla AK. G protein-coupled receptors: signaling, trafficking and regulation. In: Shukla AK, ed. *Methods in Cell Biology*. Academic Press; 2016:1 vol. 132. Online-Ressource (XXII, 480 Seiten).
54. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol*. 1996;257(2):342–358. https://doi.org/10.1006/jmbi.1996.0167.
55. Gobel U, Sander C, Schneider R, Valencia A. Correlated mutations and residue contacts in proteins. *Proteins*. 1994;18(4):309–317. https://doi.org/10.1002/prot.340180402.
56. Kaczor AA, Selent J. Oligomerization of G protein-coupled receptors: biochemical and biophysical methods. *Curr Med Chem*. 2012;18(30):4606–4634. https://doi.org/10.2174/092986711797379285.
57. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J*. 1986;5(4):823–826.
58. Dean MK, Higgs C, Smith RE, et al. Dimerization of G-protein-coupled receptors. *J Med Chem*. 2001;44(26):4595–4614. https://doi.org/10.1021/jm010290+.
59. Latek D, Trzakowski B, Niewieczerzal S, et al. Modeling of membrane proteins. In: Liwo A, ed. *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*. 2nd ed. Springer; 2019:371–451.
60. Oliveira L, Paiva ACM, Vriend G. Correlated mutation analyses on very large sequence families. *Chembiochem*. 2002;3(10):1010–1017. https://doi.org/10.1002/1439-7633(20021004)3:10<1010::AID-CBIC1010>3.0.CO;2-T.
61. Gouldson PR, Dean MK, Snell CR, Bywater RP, Gkoutos G, Reynolds CA. Lipid-facing correlated mutations and dimerization in G-protein coupled receptors. *Protein Eng Des Sel*. 2001;14(10):759–767. https://doi.org/10.1093/protein/14.10.759.
62. Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG, Lichtarge O. Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *J Biol Chem*. 2004;279(9):8126–8132. https://doi.org/10.1074/jbc.M312671200.
63. Koshi JM, Goldstein RA. Context-dependent optimal substitution matrices. *Protein Eng Des Sel*. 1995;8(7):641–645. https://doi.org/10.1093/protein/8.7.641.
64. Koshi JM, Goldstein RA. Models of natural mutations including site heterogeneity. *Proteins*. 1998;32(3):289–295. https://doi.org/10.1002/(SICI)1097-0134(19980815)32:3<289::AID-PROT4>3.0.CO;2-D.

65. Koshi JM, Mindell DP, Goldstein RA. Using physical-chemistry-based substitution models in phylogenetic analyses of HIV-1 subtypes. *Mol Biol Evol*. 1999;16(2):173–179. https://doi.org/10.1093/oxfordjournals.molbev.a026100.
66. Filizola M, Weinstein H. The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics. *FEBS J*. 2005;272(12):2926–2938. https://doi.org/10.1111/j.1742-4658.2005.04730.x.
67. Reggio PH. Computational methods in drug design: modeling G protein-coupled receptor monomers, dimers, and oligomers. *AAPS J*. 2006;8(2):E322–E336. https://doi.org/10.1007/978-0-387-76678-2_3.
68. Soyer OS, Dimmic MW, Neubig RR, Goldstein RA. Dimerization in aminergic G-protein-coupled receptors: application of a hidden-site class model of evolution. *Biochemistry*. 2003;42(49):14522–14531. https://doi.org/10.1021/bi035097r.
69. Nemoto W, Toh H. Prediction of interfaces for oligomerizations of G-protein coupled receptors. *Proteins*. 2005;58(3):644–660. https://doi.org/10.1002/prot.20332.
70. Hebert TE, Moffett S, Morello JP, et al. A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J Biol Chem*. 1996;271(27):16384–16392. https://doi.org/10.1074/jbc.271.27.16384.
71. Guo W, Shi L, Javitch JA. The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer. *J Biol Chem*. 2003;278(7):4385–4388. https://doi.org/10.1074/jbc.C200679200.
72. Lee SP, O'Dowd BF, Rajaram RD, Nguyen T, George SR. D2 dopamine receptor homodimerization is mediated by multiple sites of interaction, including an intermolecular interaction involving transmembrane domain 4. *Biochemistry*. 2003;42(37):11023–11031. https://doi.org/10.1021/bi0345539.
73. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem*. 2003;278(24):21655–21662. https://doi.org/10.1074/jbc.M302536200.
74. Nemoto W, Fukui K, Toh H. GRIP: a server for predicting interfaces for GPCR oligomerization. *J Recept Signal Transduct Res*. 2009;29(6):312–317. https://doi.org/10.3109/10799890903295143.
75. Bartuzi D, Kaczor AA, Targowska-Duda KM, Matosiuk D. Recent advances and applications of molecular docking to g protein-coupled receptors. *Molecules*. 2017;22(2):1–23. https://doi.org/10.3390/molecules22020340.
76. Kaczor AA, Selent J, Sanz F, Pastor M. Modeling complexes of transmembrane proteins: systematic analysis of protein-protein docking tools. *Mol Inform*. 2013;32(8):717–733. https://doi.org/10.1002/minf.201200150.
77. Viswanath S, Dominguez L, Foster LS, Straub JE, Elber R. Extension of a protein docking algorithm to membranes and applications to amyloid precursor protein dimerization. *Proteins*. 2015;83(12):2170–2185. https://doi.org/10.1002/prot.24934.
78. Alford RF, Koehler Leman J, Weitzner BD, et al. An integrated framework advancing membrane protein modeling and design. *PLoS Comput Biol*. 2015;11(9),e1004398. https//doi.org/10.1371/journal.pcbi.1004398.
79. Hurwitz N, Schneidman-Duhovny D, Wolfson HJ. Memdock: an α-helical membrane protein docking algorithm. *Bioinformatics*. 2016;32(16):2444–2450. https://doi.org/10.1093/bioinformatics/btw184.
80. Kaczor AA, Guixà-González R, Carriõ P, et al. Multi-component protein-protein docking based protocol with external scoring for modeling dimers of G protein-coupled receptors. *Mol Inf*. 2015;34(4):246–255. https://doi.org/10.1002/minf.201400088.
81. Kaczor AA, Jörg M, Capuano B. The dopamine D2 receptor dimer and its interaction with homobivalent antagonists: homology modeling, docking and molecular dynamics. *J Mol Model*. 2016;22(9):203. https://doi.org/10.1007/s00894-016-3065-2.

82. Borroto-Escuela DO, Rodriguez D, Romero-Fernandez W, et al. Mapping the interface of a GPCR dimer: a structural model of the A2A adenosine and D2 dopamine receptor heteromer. *Front Pharmacol*. 2018;9:829. https://doi.org/10.3389/fphar.2018.00829.
83. Wang J, Miao Y. Recent advances in computational studies of GPCR-G protein interactions. In: Donev R, ed. *Advances in Protein Chemistry and Structural Biology*. Academic Press; 2019:1st ed. vol. 116. https://doi.org/10.1016/bs.apcsb.2018.11.011.
84. Altwaijry NA, Baron M, Wright DW, Coveney PV, Townsend-Nicholson A. An ensemble-based protocol for the computational prediction of helix-helix interactions in G protein-coupled receptors using coarse-grained molecular dynamics. *J Chem Theory Comput*. 2017;13(5):2254–2270. https://doi.org/10.1021/acs.jctc.6b01246.
85. Simpson LM, Taddese B, Wall ID, Reynolds CA. Bioinformatics and molecular modelling approaches to GPCR oligomerization. *Curr Opin Pharmacol*. 2010;10(1):30–37. https://doi.org/10.1016/j.coph.2009.11.001.
86. Selent J, Kaczor AA. Oligomerization of G protein-coupled receptors: computational methods. *Curr Med Chem*. 2012;18(30):4588–4605. https://doi.org/10.2174/092986711797379320.
87. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983;4(2):187–217. https://doi.org/10.1002/jcc.540040211.
88. Vanommeslaeghe K, Hatcher E, Acharya C, et al. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem*. 2010;31(4):671–690. https://doi.org/10.1002/jcc.21367.
89. Huang J, Mackerell AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem*. 2013;34(25):2135–2145. https://doi.org/10.1002/jcc.23354.
90. Case DA, Cerutti DS, Cheatham TE, et al. *Amber17*. 2017. https://doi.org/10.13140/RG.2.2.36172.41606.
91. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B*. 2007;111(27):7812–7824. https://doi.org/10.1021/jp071097f.
92. Filizola M, Wang SX, Weinstein H. Dynamic models of G-protein coupled receptor dimers: indications of asymmetry in the rhodopsin dimer from molecular dynamics simulations in a POPC bilayer. *J Comput Aided Mol Des*. 2006;20(7–8):405–416. https://doi.org/10.1007/s10822-006-9053-3.
93. Cordomí A, Perez JJ. Structural rearrangements of rhodopsin subunits in a dimer complex: a molecular dynamics simulation study. *J Biomol Struct Dyn*. 2009;27(2):127–147. https://doi.org/10.1080/07391102.2009.10507303.
94. Neri M, Vanni S, Tavernelli I, Rothlisberger U. Role of aggregation in rhodopsin signal transduction. *Biochemistry*. 2010;49(23):4827–4832. https://doi.org/10.1021/bi100478j.
95. Periole X, Huber T, Marrink S-J, Sakmar TP. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J Am Chem Soc*. 2007;129(33):10126–10132. https://doi.org/10.1021/ja0706246.
96. Periole X, Knepp AM, Sakmar TP, Marrink SJ, Huber T. Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. *J Am Chem Soc*. 2012;134(26):10959–10965. https://doi.org/10.1021/ja303286e.
97. Baltoumas FA, Theodoropoulou MC, Hamodrakas SJ. Molecular dynamics simulations and structure-based network analysis reveal structural and functional aspects of G-protein coupled receptor dimer interactions. *J Comput Aided Mol Des*. 2016;30(6):489–512. https://doi.org/10.1007/s10822-016-9919-y.

98. Ghosh A, Sonavane U, Joshi R. Multiscale modelling to understand the self-assembly mechanism of human β2-adrenergic receptor in lipid bilayer. *Comput Biol Chem*. 2014;48:29–39. https://doi.org/10.1016/j.compbiolchem.2013.11.002.
99. Prasanna X, Chattopadhyay A, Sengupta D. Cholesterol modulates the dimer interface of the β2-adrenergic receptor via cholesterol occupancy sites. *Biophys J*. 2014;106(6):1290–1300. https://doi.org/10.1016/j.bpj.2014.02.002.
100. Mondal S, Johnston JM, Wang H, Khelashvili G, Filizola M, Weinstein H. Membrane driven spatial organization of GPCRs. *Sci Rep*. 2013;3:2909. https://doi.org/10.1038/srep02909.
101. Johnston JM, Wang H, Provasi D, Filizola M. Assessing the relative stability of dimer interfaces in G protein-coupled receptors. *PLoS Comput Biol*. 2012;8(8), e1002649. https://doi.org/10.1371/journal.pcbi.1002649.
102. Rodríguez D, Gutiérrez-de-Terán H. Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor. *Proteins*. 2012;80(8):1919–1928. https://doi.org/10.1002/prot.24099.
103. Pluhackova K, Gahbauer S, Kranz F, Wassenaar TA, Böckmann RA. Dynamic cholesterol-conditioned dimerization of the G protein coupled chemokine receptor type 4. *PLoS Comput Biol*. 2016;12(11): e1005169. https://doi.org/10.1371/journal.pcbi.1005169.
104. Provasi D, Boz MB, Johnston JM, Filizola M. Preferred supramolecular organization and dimer interfaces of opioid receptors from simulated self-association. *PLoS Comput Biol*. 2015;11(3), e1004148. https://doi.org/10.1371/journal.pcbi.1004148.
105. Provasi D, Johnston JM, Filizola M. Lessons from free energy simulations of δ-opioid receptor homodimers involving the fourth transmembrane helix. *Biochemistry*. 2010;49(31):6771–6776. https://doi.org/10.1021/bi100686t.
106. Johnston JM, Aburi M, Provasi D, et al. Making structural sense of dimerization interfaces of delta opioid receptor homodimers. *Biochemistry*. 2011;50(10):1682–1690. https://doi.org/10.1021/bi101474v.
107. Zhang X, Yuan Y, Wang L, et al. Use multiscale simulation to explore the effects of the homodimerizations between different conformation states on the activation and allosteric pathway for the μ-opioid receptor. *Phys Chem Chem Phys*. 2018;20(19): 13485–13496. https://doi.org/10.1039/c8cp02016g.
108. Meral D, Provasi D, Prada-Gracia D, et al. Molecular details of dimerization kinetics reveal negligible populations of transient μ-opioid receptor homodimers at physiological concentrations. *Sci Rep*. 2018;8(1):7705. https://doi.org/10.1038/s41598-018-26070-8.
109. Kim SK, Jacobson KA. Computational prediction of homodimerization of the A3 adenosine receptor. *J Mol Graph Model*. 2006;25(4):549–561. https://doi.org/10.1016/j.jmgm.2006.03.003.
110. Fanelli F. Dimerization of the lutropin receptor: insights from computational modeling. *Mol Cell Endocrinol*. 2007;260–262:59–64. https://doi.org/10.1016/j.mce.2005.12.054.
111. Dijkman PM, Castell OK, Goddard AD, et al. Dynamic tuneable G protein-coupled receptor monomer-dimer populations. *Nat Commun*. 2018;9(1):1710. https://doi.org/10.1038/s41467-018-03727-6.
112. Erol I, Cosut B, Durdagi S. Toward understanding the impact of dimerization interfaces in angiotensin II type 1 receptor. *J Chem Inf Model*. 2019;59(10):4314–4327. https://doi.org/10.1021/acs.jcim.9b00294.
113. Liu X, Kai M, Jin L, Wang R. Computational study of the heterodimerization between μ and δ receptors. *J Comput Aided Mol Des*. 2009;23(6):321–332. https://doi.org/10.1007/s10822-009-9262-7.

114. Wang L, Yuan Y, Chen X, et al. Probing the cooperative mechanism of the μ-δ opioid receptor heterodimer by multiscale simulation. *Phys Chem Chem Phys*. 2018;20(47): 29969–29982. https://doi.org/10.1039/c8cp06652c.
115. Navarro G, Cordomí A, Zelman-Femiak M, et al. Quaternary structure of a G-protein-coupled receptor heterotetramer in complex with Gi and Gs. *BMC Biol*. 2016;14(1):26. https://doi.org/10.1186/s12915-016-0247-4.
116. Bruno A, Guadix AE, Costantino G. Molecular dynamics simulation of the heterodimeric mGluR2/5HT$_{2A}$ complex. An atomistic resolution study of a potential new target in psychiatric conditions. *J Chem Inf Model*. 2009;49(6):1602–1616. https://doi.org/10.1021/ci900067g.
117. Witt M, Ślusarz MJ, Ciarkowski J. Molecular modeling of vasopressin V2 receptor tetramer in hydrated lipid membrane. *QSAR Comb Sci*. 2008;27(6):684–693. https://doi.org/10.1002/qsar.200730082.
118. Greenwood D. An overview of neural networks. *Behav Sci*. 1991;36(1):1–33. https://doi.org/10.1002/bs.3830360102.
119. Ding S, Li H, Su C, Yu J, Jin F. Evolutionary artificial neural networks: a review. *Artif Intell Rev*. 2013;39(3):251–260. https://doi.org/10.1007/s10462-011-9270-6.
120. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.
121. Ali J, Khan R, Ahmad N, Maqsood I. Random forests and decision trees. *Int J Comput Sci Issues*. 2012;9(5):272–278, www.IJCSI.org. [(Accessed 29 October 2019)].
122. Polikar R. Ensemble based systems in decision making. *IEEE Circuits Syst Mag*. 2006;6(3):21–44. https://doi.org/10.1109/MCAS.2006.1688199.
123. Moreira IS, Koukos PI, Melo R, et al. SpotOn: high accuracy identification of protein-protein interface hot-spots. *Sci Rep*. 2017;7(1):8007. https://doi.org/10.1038/s41598-017-08321-2.
124. Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*. 2005;39(1):309–338. https://doi.org/10.1146/annurev.genet.39.073003.114725.
125. Dey A. Machine learning algorithms: a review. *Int J Comput Sci Inf Technol*. 2016; 7(3):1174–1179, www.ijcsit.com. [(Accessed 29 October 2019)].
126. Navío D, Rosell M, Aguirre J, de la Cruz X, Fernández-Recio J. Structural and computational characterization of disease-related mutations involved in protein-protein interfaces. *Int J Mol Sci*. 2019;20(7). E1583. https://doi.org/10.3390/ijms20071583.
127. Yuan Z, Zhang F, Davis MJ, Bodén M, Teasdale RD. Predicting the solvent accessibility of transmembrane residues from protein sequence. *J Proteome Res*. 2006;5(5):1063–1070. https://doi.org/10.1021/pr050397b.
128. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov*. 1998;2(2):121–167. https://doi.org/10.1023/A:1009715923555.
129. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput*. 2004;14(3):199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.
130. Koppisetti RK, Fulcher YG, Jurkevich A, et al. Ambidextrous binding of cell and membrane bilayers by soluble matrix metalloproteinase-12. *Nat Commun*. 2014;5:5552. https://doi.org/10.1038/ncomms6552.
131. Jones S, Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci USA*. 1996;93:13–20. https://doi.org/10.1073/pnas.93.1.13.
132. Xue LC, Dobbs D, Bonvin AMJJ, Honavar V. Computational prediction of protein interfaces: a review of data driven methods. *FEBS Lett*. 2015;589(23):3516–3526. https://doi.org/10.1016/j.febslet.2015.10.003.
133. Pons C, Glaser F, Fernandez-Recio J. Prediction of protein-binding areas by small-world residue networks and application to docking. *BMC Bioinformatics*. 2011;12: 378. https://doi.org/10.1186/1471-2105-12-378.

134. Jordan RA, El-Manzalawy Y, Dobbs D, Honavar V. Predicting protein-protein interface residues using local surface structural similarity. *BMC Bioinformatics*. 2012;13(1):41. https://doi.org/10.1186/1471-2105-13-41.
135. Murakami Y, Mizuguchi K. Applying the Naïve Bayes classifier with kernel density estimation to the prediction of protein-protein interaction sites. *Bioinformatics*. 2010;26(15):1841–1848. https://doi.org/10.1093/bioinformatics/btq302.
136. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins*. 2007;66(3):630–645. https://doi.org/10.1002/prot.21248.
137. Savojardo C, Fariselli P, Martelli PL, Casadio R. ISPRED4: interaction sites PREDiction in protein structures with a refining grammar model. *Bioinformatics*. 2017;33(11):1656–1663. https://doi.org/10.1093/bioinformatics/btx044.
138. Hou Q, De Geest PFG, Griffioen CJ, Abeln S, Heringa J, Feenstra KA. SeRenDIP: SEquential REmasteriNg to DerIve profiles for fast and accurate predictions of PPI interface positions. *Bioinformatics*. 2019;35(22):4794–4796. https://doi.org/10.1093/bioinformatics/btz428.
139. Sanchez-Garcia R, Sorzano COS, Carazo JM, Segura J. BIPSPI: a method for the prediction of partner-specific protein-protein interfaces. *Bioinformatics*. 2019;35(3):470–477. https://doi.org/10.1093/bioinformatics/bty647.
140. Zeng H, Wang S, Zhou T, et al. ComplexContact: a web server for inter-protein contact prediction using deep learning. *Nucleic Acids Res*. 2018;46(W1):W432–W437. https://doi.org/10.1093/nar/gky420.
141. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. *Proteins*. 2005;61(1):21–35. https://doi.org/10.1002/prot.20514.
142. de Vries SJ, Bonvin AMJJ. Cport: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS One*. 2011;6(3): e17695. https://doi.org/10.1371/journal.pone.0017695.
143. Hashemifar S, Neyshabur B, Khan AA, Xu J. Predicting protein-protein interactions through sequence-based deep learning. *Bioinformatics*. 2018;34(17):i802–i810. https://doi.org/10.1093/bioinformatics/bty573.
144. Meyer MJ, Beltrán JF, Liang S, et al. Interactome INSIDER: a structural interactome browser for genomic studies. *Nat Methods*. 2018;15(2):107–114. https://doi.org/10.1038/nmeth.4540.
145. Hopf TA, Schärfe CPI, Rodrigues JPGLM, et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife*. 2014;3: e03430. https://doi.org/10.7554/eLife.03430.
146. Negi SS, Schein CH, Oezguen N, Power TD, Braun W. InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics*. 2007;23(24):3397–3399. https://doi.org/10.1093/bioinformatics/btm474.
147. Qin S, Zhou HX. Meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*. 2007;23(24):3386–3387. https://doi.org/10.1093/bioinformatics/btm434.
148. Afsar Minhas F, Geiss BJ, Ben-Hur A. PAIRpred: partner-specific prediction of interacting residues from sequence and structure. *Proteins*. 2014;82(7):1142–1155. https://doi.org/10.1002/prot.24479.
149. Kufareva I, Budagyan L, Raush E, Totrov M, Abagyan R. PIER: protein interface recognition for structural proteomics. *Proteins*. 2007;67(2):400–417. https://doi.org/10.1002/prot.21233.
150. Liang S, Zhang C, Liu S, Zhou Y. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res*. 2006;34(13):3698–3707. https://doi.org/10.1093/nar/gkl454.

151. Ahmad S, Mizuguchi K. Partner-aware prediction of interacting residues in protein-protein complexes from sequence data. *PLoS One*. 2011;6(12):29104. https://doi.org/10.1371/journal.pone.0029104.
152. Chatterjee P, Basu S, Kundu M, Nasipuri M, Plewczynski D. PPI_SVM: prediction of protein-protein interactions using machine learning, domain-domain affinities and frequency tables. *Cell Mol Biol Lett*. 2011;16(2):264–278. https://doi.org/10.2478/s11658-011-0008-x.
153. Kuo TH, Bin LK. Predicting protein-protein interaction sites using sequence descriptors and site propensity of neighboring amino acids. *Int J Mol Sci*. 2016;17(11):1788. https://doi.org/10.3390/ijms17111788.
154. Zellner H, Staudigel M, Trenner T, et al. Prescont: predicting protein-protein interfaces utilizing four residue properties. *Proteins*. 2012;80(1):154–168. https://doi.org/10.1002/prot.23172.
155. Zhang QC, Deng L, Fisher M, Guan J, Honig B, Petrey D. PredUs: a web server for predicting protein interfaces using structural neighbors. *Nucleic Acids Res*. 2011;39:W283–W287. https://doi.org/10.1093/nar/gkr311.
156. Baspinar A, Cukuroglu E, Nussinov R, Keskin O, Gursoy A. PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. *Nucleic Acids Res*. 2014;42:W285–W289. https://doi.org/10.1093/nar/gku397.
157. Ofran Y, Rost B. ISIS: interaction sites identified from sequence. *Bioinformatics*. 2007;23(2):13–16. https://doi.org/10.1093/bioinformatics/btl303.
158. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol*. 2004;338(1):181–199. https://doi.org/10.1016/j.jmb.2004.02.040.
159. Murakami Y, Jones S. SHARP 2: protein-protein interaction predictions using patch analysis. *Bioinformatics*. 2006;22(14):1794–1795. https://doi.org/10.1093/bioinformatics/btl171.
160. Valente GT, Acencio ML, Martins C, Lemke N. The development of a universal in silico predictor of protein-protein interactions. *PLoS One*. 2013;8(5):65587. https://doi.org/10.1371/journal.pone.0065587.
161. De Vries SJ, Van Dijk ADJ, AMJJ B. WHISCY: what information does surface conservation yield? Application to data-driven docking. *Proteins*. 2006;63(3):479–489. https://doi.org/10.1002/prot.20842.
162. Alonso-López Di C-LFJ, Gutiérrez MA, et al. APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database*. 2019;2019:5. https://doi.org/10.1093/database/baz005.
163. Oughtred R, Stark C, Breitkreutz BJ, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D529–D541. https://doi.org/10.1093/nar/gky1079.
164. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*. 2013;41(D1):D793–D800. https://doi.org/10.1093/nar/gks1055.
165. Salwinski L. The database of interacting proteins: 2004 update. *Nucleic Acids Res*. 2004;32(90001):449D–451. https://doi.org/10.1093/nar/gkh086.
166. Munk C, Isberg V, Mordalski S, et al. GPCRdb: the G protein-coupled receptor database—an introduction. *Br J Pharmacol*. 2016;173(14):2195–2207. https://doi.org/10.1111/bph.13509.
167. Theodoropoulou MC, Bagos PG, Spyropoulos IC, Hamodrakas SJ. gpDB: a database of GPCRs, G-proteins, effectors and their interactions. *Bioinformatics*. 2008;24(12):1471–1472. https://doi.org/10.1093/bioinformatics/btn206.

168. Keshava Prasad TS, Goel R, Kandasamy K, et al. Human protein reference database—2009 update. *Nucleic Acids Res*. 2009;37(Suppl. 1):D767–D772. https://doi.org/10.1093/nar/gkn892.
169. Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: Context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res*. 2019;47(D1):D581–D589. https://doi.org/10.1093/nar/gky1037.
170. Breuer K, Foroushani AK, Laird MR, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res*. 2013;41(D1):D1228–D1233. https://doi.org/10.1093/nar/gks1147.
171. Kerrien S, Aranda B, Breuza L, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. 2012;40(D1):D841–D846. https://doi.org/10.1093/nar/gkr1088.
172. Mosca R, Céol A, Aloy P. Interactome3D: adding structural details to protein networks. *Nat Methods*. 2013;10(1):47–53. https://doi.org/10.1038/nmeth.2289.
173. Turner B, Razick S, Turinsky AL, et al. iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford)*. 2010;2010:1–15. https://doi.org/10.1093/database/baq023.
174. Calderone A, Castagnoli L, Cesareni G. Mentha: a resource for browsing integrated protein-interaction networks. *Nat Methods*. 2013;10(8):690–691. https://doi.org/10.1038/nmeth.2561.
175. Licata L, Briganti L, Peluso D, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res*. 2012;40(D1):D857–D861. https://doi.org/10.1093/nar/gkr930.
176. Pagel P, Kovac S, Oesterheld M, et al. The MIPS mammalian protein-protein interaction database. *Bioinformatics*. 2005;21(6):832–834. https://doi.org/10.1093/bioinformatics/bti115.
177. Blohm P, Frishman G, Smialowski P, et al. Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res*. 2014;42(D1):D396–D400. https://doi.org/10.1093/nar/gkt1079.
178. Gioutlakis A, Klapa MI, Moschonas NK. PICKLE 2.0: a human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLoS One*. 2017;12(10), e0186039. https://doi.org/10.1371/journal.pone.0186039.
179. Cowley MJ, Pinese M, Kassahn KS, et al. PINA v2.0: mining interactome modules. *Nucleic Acids Res*. 2012;40(D1):D862–D865. https://doi.org/10.1093/nar/gkr967.
180. Zhang QC, Petrey D, Garzón JI, Deng L, Honig B. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res*. 2013;41(D1):D828–D833. https://doi.org/10.1093/nar/gks1231.
181. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–D613. https://doi.org/10.1093/nar/gky1131.
182. Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Brief Bioinform*. 2009;10(3):217–232. https://doi.org/10.1093/bib/bbp001.
183. Xue LC, Dobbs D, Honavar V. HomPPI: a class of sequence homology based protein-protein interface prediction methods. *BMC Bioinform*. 2011;12:244. https://doi.org/10.1186/1471-2105-12-244.
184. Liu Q, Li Z, Li J. Use B-factor related features for accurate classification between protein binding interfaces and crystal packing contacts. *BMC Bioinform*. 2014;15(16). S3. https://doi.org/10.1186/1471-2105-15-S16-S3.
185. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577–2637. https://doi.org/10.1002/bip.360221211.

186. Moreira IS, Fernandes PA, Ramos MJ. Hot spots—a review of the protein-protein interface determinant amino-acid residues. *Proteins*. 2007;68(4):803–812. https://doi.org/10.1002/prot.21396.
187. Zhu X, Mitchell JC. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins*. 2011;79(9):2671–2683. https://doi.org/10.1002/prot.23094.
188. Wang H, Liu C, Deng L. Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting. *Sci Rep*. 2018;8(1):14285. https://doi.org/10.1038/s41598-018-32511-1.
189. Lise S, Buchan D, Pontil M, Jones DT. Predictions of hot spot residues at protein-protein interfaces using support vector machines. *PLoS One*. 2011;6(2):16774. https://doi.org/10.1371/journal.pone.0016774.
190. Munteanu CR, Pimenta AC, Fernandez-Lozano C, Melo A, Cordeiro MNDS, Moreira IS. Solvent accessible surface area-based hot-spot detection methods for protein-protein and protein-nucleic acid interfaces. *J Chem Inf Model*. 2015;55(5):1077–1086. https://doi.org/10.1021/ci500760m.
191. Koes DR, Camacho CJ. PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. *Nucleic Acids Res*. 2012;40(W1):W387–W392. https://doi.org/10.1093/nar/gks336.
192. Park Y, Hayat S, Helms V. Prediction of the burial status of transmembrane residues of helical membrane proteins. *BMC Bioinform*. 2007;8:302. https://doi.org/10.1186/1471-2105-8-302.
193. Hayat M, Khan A. WRF-TMH: predicting transmembrane helix by fusing composition index and physicochemical properties of amino acids. *Amino Acids*. 2013;44(5):1317–1328. https://doi.org/10.1007/s00726-013-1466-4.
194. Yin X, Yang J, Xiao F, Yang Y, Bin SH. MemBrain: an easy-to-use online webserver for transmembrane protein structure prediction. *Nano Micro Lett*. 2018;10(1):2. https://doi.org/10.1007/s40820-017-0156-2.
195. Adamian L, Liang J. Prediction of transmembrane helix orientation in polytopic membrane proteins. *BMC Struct Biol*. 2006;6:13. https://doi.org/10.1186/1472-6807-6-13.
196. Nugent T, Ward S, Jones DT. The MEMPACK alpha-helical transmembrane protein structure prediction server. *Bioinformatics*. 2011;27(10):1438–1439. https://doi.org/10.1093/bioinformatics/btr096.
197. Lo A, Chiu YY, Rødland EA, Lyu PC, Sung TY, Hsu WL. Predicting helix-helix interactions from residue contacts in membrane proteins. *Bioinformatics*. 2009;25(8):996–1003. https://doi.org/10.1093/bioinformatics/btp114.
198. Fuchs A, Kirschner A, Frishman D. Prediction of helix-helix contacts and interacting helices in polytopic membrane proteins using neural networks. *Proteins*. 2009;74(4):857–871. https://doi.org/10.1002/prot.22194.
199. Wang XF, Chen Z, Wang C, Yan RX, Zhang Z, Song J. Predicting residue-residue contacts and helix-helix interactions in transmembrane proteins using an integrative feature-based random forest approach. *PLoS One*. 2011;6(10):26767. https://doi.org/10.1371/journal.pone.0026767.
200. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct*. 2012;7:12. https://doi.org/10.1186/1745-6150-7-12.
201. Mancusso R, Gregorio GG, Liu Q, Wang DN. Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. *Nature*. 2012;491(7425):622–626. https://doi.org/10.1038/nature11542.
202. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44:W344–W350. https://doi.org/10.1093/nar/gkw408.

203. Taylor MS, Fung HK, Rajgaria R, Filizola M, Weinstein H, Floudas CA. Mutations affecting the oligomerization interface of G-protein-coupled receptors revealed by a novel de novo protein design framework. *Biophys J*. 2008;94(7):2470–2481. https://doi.org/10.1529/biophysj.107.117622.
204. Popov P, Peng Y, Shen L, et al. Computational design of thermostabilizing point mutations for G protein-coupled receptors. *eLife*. 2018;7: e34729. https://doi.org/10.7554/eLife.34729.
205. Nicoludis JM, Gaudet R. Applications of sequence coevolution in membrane protein biochemistry. *Biochim Biophys Acta Biomembr*. 2018;1860(4):895–908. https://doi.org/10.1016/j.bbamem.2017.10.004.
206. Marks DS, Colwell LJ, Sheridan R, et al. Protein 3D structure computed from evolutionary sequence variation. *PLoS One*. 2011;6(12):28766. https://doi.org/10.1371/journal.pone.0028766.
207. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–D285. https://doi.org/10.1093/nar/gkv1344.
208. Bitbol AF, Dwyer RS, Colwell LJ, Wingreen NS. Inferring interaction partners from protein sequences. *Proc Natl Acad Sci USA*. 2016;113(43):12180–12185. https://doi.org/10.1073/pnas.1606762113.
209. Gueudré T, Baldassi C, Zamparo M, Weigt M, Pagnani A. Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis. *Proc Natl Acad Sci USA*. 2016;113(43):12186–12191. https://doi.org/10.1073/pnas.1607570113.
210. Szurmant H, Weigt M. Inter-residue, inter-protein and inter-family coevolution: bridging the scales. *Curr Opin Struct Biol*. 2018;50:26–32. https://doi.org/10.1016/j.sbi.2017.10.014.
211. Wu Q, Peng Z, Anishchenko I, Cong Q, Baker D, Yang J. Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics*. 2019; https://doi.org/10.1093/bioinformatics/btz477.
212. Bordner AJ. Predicting protein-protein binding sites in membrane proteins. *BMC Bioinformatics*. 2009;10:312. https://doi.org/10.1186/1471-2105-10-312.
213. Zeng B, Hönigschmid P, Frishman D. Residue co-evolution helps predict interaction sites in α-helical membrane proteins. *J Struct Biol*. 2019;206(2):156–169. https://doi.org/10.1016/j.jsb.2019.02.009.
214. Asadabadi EB, Abdolmaleki P. Predictions of protein-protein interfaces within membrane protein complexes. *Avicenna J Med Biotechnol*. 2013;5(3):148–157.
215. Li B, Mendenhall J, Meiler J. Interfaces between alpha-helical integral membrane proteins: characterization, prediction, and docking. *Comput Struct Biotechnol J*. 2019;17:699–711. https://doi.org/10.1016/j.csbj.2019.05.005.
216. Li B, Mendenhall J, Nguyen ED, Weiner BE, Fischer AW, Meiler J. Accurate prediction of contact numbers for multi-spanning helical membrane proteins. *J Chem Inf Model*. 2016;56(2):423–434. https://doi.org/10.1021/acs.jcim.5b00517.
217. Li B, Rong D, Wang Y. Targeting protein-protein interaction with covalent small-molecule inhibitors. *Curr Top Med Chem*. 2019;19(21):1872–1876. https://doi.org/10.2174/1568026619666191011163410.
218. Ni D, Lu S, Zhang J. Emerging roles of allosteric modulators in the regulation of protein-protein interactions (PPIs): a new paradigm for PPI drug discovery. *Med Res Rev*. 2019;39(6):2314–2342. https://doi.org/10.1002/med.21585.
219. Ni D, Liu N, Sheng C. Allosteric modulators of protein-protein interactions (PPIs). In: Zhang J, Nussinov R, eds. *Protein Allostery in Drug Discovery*. Singapore: Springer; 2019: Advances in Experimental Medicine and Biology. vol. 1163.
220. Jin L, Wang W, Fang G. Targeting protein-protein interaction by small molecules. *Annu Rev Pharmacol Toxicol*. 2014;54:435–456. https://doi.org/10.1146/annurev-pharmtox-011613-140028.

221. Nichols SE, Hernández CX, Wang Y, McCammon JA. Structure-based network analysis of an evolved G protein-coupled receptor homodimer interface. *Protein Sci*. 2013;22(6):745–754. https://doi.org/10.1002/pro.2258.
222. Cukuroglu E, Engin HB, Gursoy A, Keskin O. Hot spots in protein-protein interfaces: towards drug discovery. *Prog Biophys Mol Biol*. 2014;116(2–3):165–173. https://doi.org/10.1016/j.pbiomolbio.2014.06.003.
223. Jubb H, Blundell TL, Ascher DB. Flexibility and small pockets at protein-protein interfaces: new insights into druggability. *Prog Biophys Mol Biol*. 2015;119(1):2–9. https://doi.org/10.1016/j.pbiomolbio.2015.01.009.
224. London N, Raveh B, Schueler-Furman O. Druggable protein-protein interactions—from hot spots to hot segments. *Curr Opin Chem Biol*. 2013;17(6):952–959. https://doi.org/10.1016/j.cbpa.2013.10.011.
225. Guo W, Wisniewski JA, Ji H. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett*. 2014;24(11):2546–2554. https://doi.org/10.1016/j.bmcl.2014.03.095.
226. Rosell M, Fernandez-Recio J. Hot-spot analysis for drug discovery targeting protein-protein interactions. *Expert Opin Drug Discovery*. 2018;13(4):327–338. https://doi.org/10.1080/17460441.2018.1430763.
227. Vellani V, Giacomoni C. Gabapentin inhibits protein kinase C epsilon translocation in cultured sensory neurons with additive effects when coapplied with paracetamol (acetaminophen). *ScientificWorldJournal*. 2017;2017:3595903. https://doi.org/10.1155/2017/3595903.
228. Wang Y, Peng Y, Zhang B, et al. Targeting trimeric transmembrane domain 5 of oncogenic latent membrane protein 1 using a computationally designed peptide. *Chem Sci*. 2019;10(32):7584–7590. https://doi.org/10.1039/c9sc02474c.
229. Mujic-Delic A, de Wit RH, Verkaar F, Smit MJ. GPCR-targeting nanobodies: attractive research tools, diagnostics, and therapeutics. *Trends Pharmacol Sci*. 2014;35(5):247–255. https://doi.org/10.1016/j.tips.2014.03.003.
230. Jahnichen S, Blanchetot C, Maussang D, et al. CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells. *Proc Natl Acad Sci USA*. 2010;107(47):20565–20570. https://doi.org/10.1073/pnas.1012865107.
231. Jastrzebska B, Chen Y, Orban T, Jin H, Hofmann L, Palczewski K. Disruption of rhodopsin dimerization with synthetic peptides targeting an interaction interface. *J Biol Chem*. 2015;290(42):25728–25744. https://doi.org/10.1074/jbc.M115.662684.
232. Plach M, Schafer T, Borroto-Escuela DO, et al. Differential allosteric modulation within dopamine D2R—neurotensin NTS1R and D2R—serotonin 5-HT2AR receptor complexes gives bias to intracellular calcium signalling. *Sci Rep*. 2019;9(1):16312. https://doi.org/10.1038/s41598-019-52540-8.
233. Higueruelo AP, Schreyer A, Bickerton GRJ, Pitt WR, Groom CR, Blundell TL. Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem Biol Drug Des*. 2009;74(5):457–467. https://doi.org/10.1111/j.1747-0285.2009.00889.x.
234. Basse MJ, Betzi S, Bourgeas R, et al. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res*. 2013;41(Database issue):D824–D827. https://doi.org/10.1093/nar/gks1002.
235. Labbé CM, Laconde G, Kuenemann MA, Villoutreix BO, Sperandio O. IPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug Discov Today*. 2013;18(19–20):958–968. https://doi.org/10.1016/j.drudis.2013.05.003.
236. Labbé CM, Kuenemann MA, Zarzycka B, et al. IPPI-DB: an online database of modulators of protein-protein interactions. *Nucleic Acids Res*. 2016;44(D1):D542–D547. https://doi.org/10.1093/nar/gkv982.