Advances in Computational Techniques to Study GPCR–Ligand Recognition

Antonella Ciancetta,[1] Davide Sabbadin,[1] Stephanie Federico,[2] Giampiero Spalluto,[2] and Stefano Moro[1,*]

G-protein-coupled receptors (GPCRs) are among the most intensely investigated drug targets. The recent revolutions in protein engineering and molecular modeling algorithms have overturned the research paradigm in the GPCR field. While the numerous ligand-bound X-ray structures determined have provided invaluable insights into GPCR structure and function, the development of algorithms exploiting graphics processing units (GPUs) has made the simulation of GPCRs in explicit lipid–water environments feasible within reasonable computation times. In this review we present a survey of the recent advances in structure-based drug design approaches with a particular emphasis on the elucidation of the ligand recognition process in class A GPCRs by means of membrane molecular dynamics (MD) simulations.

Trends

GPCRs are among the most intensely investigated drug targets.

Recent numerous ligand-bound X-ray structures have provided invaluable insights into GPCR structure and function.

Structural modeling of GPCR activation/inactivation promises to accelerate drug discovery in this field.

A Long and Winding Road

Research focused on GPCRs experienced an extraordinary event in 2012: Robert Lefkowitz and Brian Kobilka were awarded the Nobel Prize in Chemistry for their pioneering discoveries in this field. In 1968, Lefkowitz used radioactivity to locate cell receptors by tracking the radiation of an iodine isotope attached to several hormones and identified a receptor for adrenalin [1]. In the 1980s, Kobilka joined Lefkowitz's team and isolated the gene that encodes the β-adrenoceptor from the human genome [2]. From their genetic studies, they demonstrated that there is a whole family of receptors that look alike and function in the same manner. Furthermore, in 2007, Kobilka and his research team achieved another fundamental breakthrough by solving the crystallographic structure of the β_2-adrenoceptor [3].

Today it is widely recognized that GPCRs represent the largest family of surface receptors, with more than 800 members in humans [4]. They respond to various extracellular (EC) stimuli ranging from small molecules to lipids, peptides, proteins, and even light [5]. The binding event triggers the activation of at least two distinct classes of signaling partners [6], such as β-arrestins and cytoplasmic heterotrimeric GTP-binding proteins (G proteins), and mediates signal transduction through the modulation of several downstream effectors. The participation of GPCRs in numerous physiopathological processes entails a potential role for their modulation by agonists, antagonists, and inverse agonists in the treatment of several diseases, including cardiovascular and mental disorders [7], cancer [8], and viral infections [9]. Recently, targeting GPCRs through allosteric binders has emerged as a strategy to design drugs to treat disorders affecting the central nervous system (CNS) [10]. Currently, approximately 50% of drugs in clinical use exert their effects by acting on GPCR-mediated signaling pathways [11].

[1]Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università Padova, via Marzolo 5, I-35131 Padova, Italy
[2]Dipartimento di Scienze Chimiche e Farmaceutiche, Università di Trieste, Piazzale Europa 1, I-34127 Trieste, Italy

*Correspondence: stefano.moro@unipd.it (S. Moro).

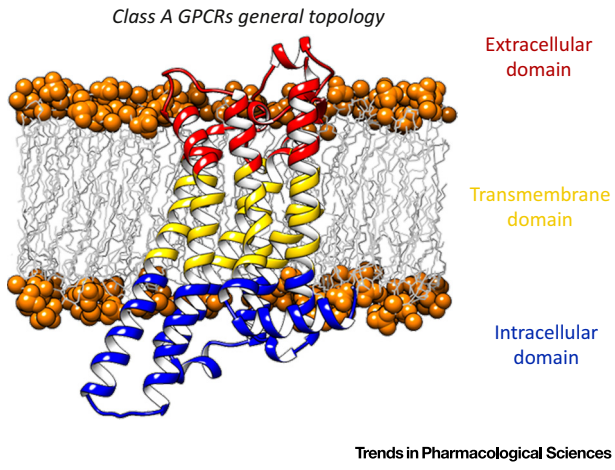


Figure 1. Overview of the Distinct Functional Regions in Class A GPCRs: the Extracellular (EC) Domain; the Transmembrane (TM) Domain; and the Intracellular (IC) Domain.

Glossary

Homology modeling: computational technique aimed at predicting the 3D structure of a protein starting from its primary sequence. The method uses the experimentally determined (by X-rays or, less frequently, by NMR) structure of a homologous protein as a template to model the target structure. The approach relies on the alignment between the template and the target sequences. Although it is generally accepted that 3D structures are more conserved than primary sequences, it is good practice to apply the method only if the sequence identity is above the 30% threshold.

Molecular docking: computational technique aimed at predicting the most favorable 3D conformation of a ligand–protein complex. In the approach most commonly applied, the protein structure is held fixed while the conformational space of the ligand is explored by a search algorithm. A key element of any docking algorithm is the scoring function; that is, an equation that ranks the generated ligand–protein conformations by assigning them a fitness value (an adimensional score, the higher the better) or a pseudo-energetic value (an estimate of the binding free energy, the lower the better).

Molecular dynamics: computational technique aimed at predicting the temporal evolution of biomolecular systems by solving Newton's equation of motion.

Virtual screening: computational approach used to identify chemical structures that are predicted to have particular properties. For example, in the context of drug discovery, it may involve computationally searching large libraries of chemical structures to identify those structures that are most likely to bind to a drug target.

According to the GRAFS classification [12], human GPCRs are commonly grouped into five main classes: glutamate (class C), rhodopsin (class A), adhesion (class B), secretin (class B), and frizzled/taste2 (class F). From a structural viewpoint, all members share a common architecture represented by seven membrane-spanning helices connected by three intracellular and three EC loops (ICLs and ECLs, respectively) with the N-terminal domain exposed toward the EC side. In class A GPCRs, three distinct functional regions can be identified (Figure 1): (i) the EC domain [N-terminal domain, ECLs, and upper transmembrane (TM) region], where ligand recognition and binding occur; (ii) the TM domain (TM bundle core), which experiences the largest conformational changes on ligand binding; and (iii) the intracellular domain (lower TM region, ICLs, and C-terminal domain), which couples to signaling partners.

The insertion of GPCRs into the cell membrane, along with the receptors' dynamism, initially hampered the structural determination of GPCRs. Recent advances in protein engineering and crystallography have represented a breakthrough and yielded numerous X-ray structures [13,14]. In particular, advanced techniques such as the engineering of chimera receptors using fusion proteins [15], receptor complexation with antibody fragments [16], or their thermostabilization through systematic scanning mutagenesis [17] have allowed researchers to overcome the limitations on GPCR structural determination. The availability of numerous ligand-bound 3D structures provides invaluable insights to understand GPCR function and pharmacology and enables the application of structure-based drug design (SBDD) approaches to aid the discovery of novel candidates with improved pharmacological profiles (Figure 2) [18–21]. In addition, the recent exploitation of commodity GPUs, a technology first designed to improve video game performance, in drug design (GPU-driven drug design) represents another important step forward for several drug discovery fields [22]. In particular, running calculations on GPCRs using GPUs enables molecular dynamics simulations (see Glossary) in explicit lipid–water environment within reasonable computation times [23–25]. Therefore, MD has become a helpful complement for the study of GPCR biophysics and molecular pharmacology by enriching our understanding of, among other aspects, ligand–receptor interaction [26] and ligand–subtype selectivity [27]. Until recently, lengthy simulations of GPCRs required specialized supercomputers and thus were the prerogative of only a few research groups. However, the emerging exploitation of cloud computing [28] as well as the development of new MD-based algorithms will allow a wider community to achieve long-timescale simulations in the near future.

In this review, we briefly survey the recent methodological advances in the field of ligand–GPCR recognition process simulations. For simplicity, we limit the discussion to class A GPCRs, the

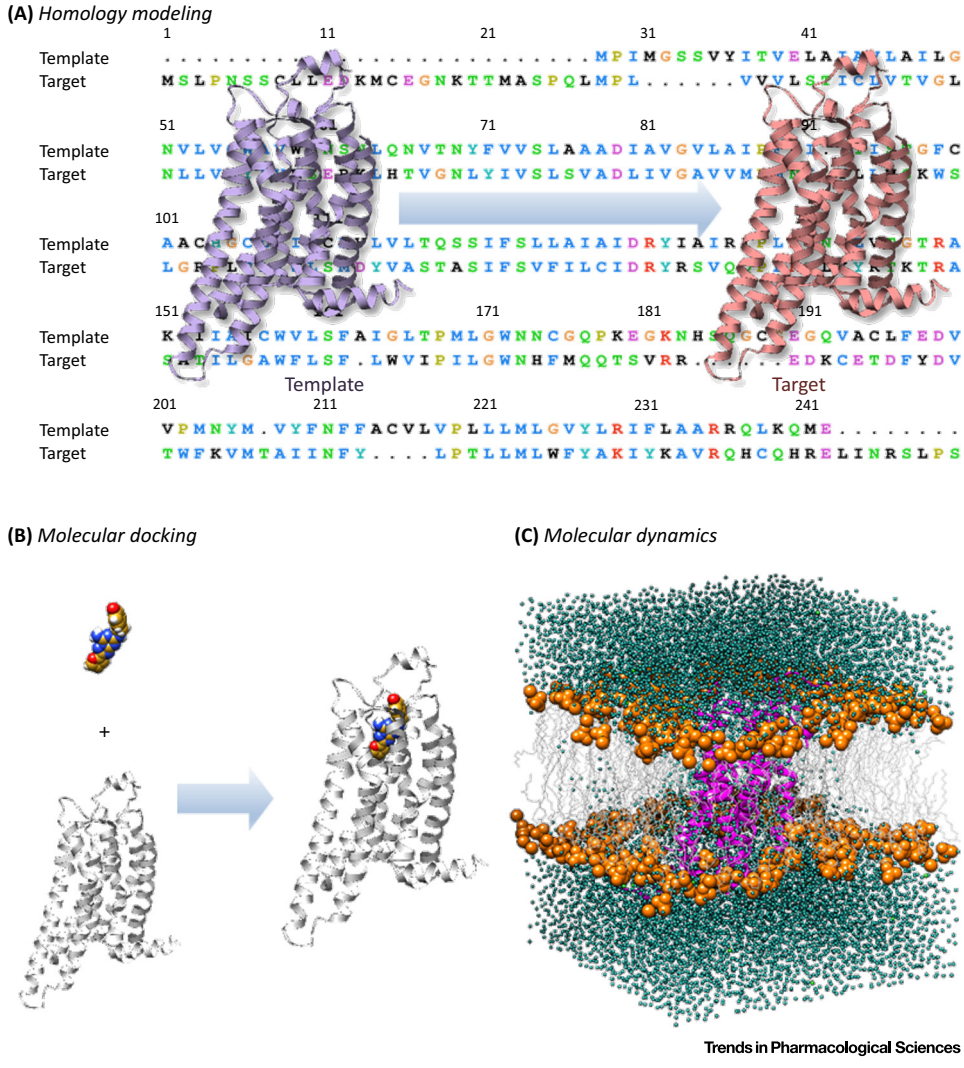


Figure 2. Overview of the Structure-Based Drug Design Techniques Discussed in this Review. (A) The construction of receptor 3D structure through homology modeling; (B) the prediction of ligand–receptor 3D conformation through molecular docking; (C) the study of ligand–receptor dynamics in an explicit lipid–water environment.

most widely investigated and experimentally characterized GPCR family. In the following, the description of the more recent methodological advances is discussed along with a few examples of their applications in solving 'real-life problems' in drug discovery taken from recent literature.

Expanding Structural Information

The GPCR 3D structures that have become recently available represent invaluable starting points for structure-based drug design approaches [29]. To date, the Protein Data Bank (PDB) contains approximately 100 ligand-bound GPCRs solved by X-rays. These structures, however, cover only ~1% of the GPCR family members [30]. Therefore, research focused on the majority of the drug targets of interest relies on homology models. In principle, any of the available X-ray structures can serve as a template to gather structural information on homologous GPCRs, offering the possibility to spread the 3D knowledge over other family members (Figure 2A). Although the sequence identity among homologous proteins is often below the recommended

30% threshold, the structural constraints imposed by the transmembrane helical domains as well as the presence of highly conserved structural motifs makes this approach generally successful [31,32]. Consequently, in the past years, in parallel to the rapid growth of structural information, an increasing number of *ad hoc* web services focused on GPCRs have been released to the scientific community [33–36]. These resources provide users with the ability to download pregenerated models and/or to build their own homology models. In particular, in the latest update of the GPCRdb, homology models for all class A receptors are available. To derive them, the selection of the templates was based on structure-based alignments in agreement with the newly available structural data. When multiple structures of a suitable template in complex with ligands of different types, receptor-binding affinities, and molecular weights are available – as in the case of β-adrenoceptors and the human adenosine A_{2A} receptor (hA_{2A} AR) – the selection of a proper template is not straightforward. We have recently proposed a possible strategy to guide the selection by implementing the Best Template Searching tool [37] in our Adenosiland platform [35], a web resource dedicated to adenosine receptors (ARs). Through this functionality, a template is suggested according to the similarity of a query structure with the ligands cocrystallized with the hA_{2A} AR subtype. Although focused on ARs, the underlying philosophy can be applied to other class A GPCRs that have been cocrystallized with different ligands. The AR models available on the platform were refined with 20 ns of membrane MD simulations and represent alternative structural starting points to the conventional homology models. Interestingly, the ability to refine homology models through 2.5 ns of membrane MD simulations has also been recently implemented in the GPCR-ModSim platform [38].

Reproducing or Anticipating X-Ray Binding Modes

Once a 3D structure (gathered experimentally or through **homology modeling**) has been achieved, several SBDD techniques can be applied to inspect ligand–GPCR binding. Among these, **molecular docking** (Figure 2B) is one of the most widely used, due also to the modest amount of time and computational resources required to perform it. When judiciously applied, docking can represent a helpful tool to rationalize the structure–activity relationships (SARs) of known binders or to scout novel candidates in **virtual screening** (VS) campaigns. Moreover, in the GPCR research field, it has been convincingly demonstrated that docking-based VS can be an effective strategy to identify novel hits [39–43]. As is generally acknowledged, the success rate of docking-based approaches strongly depends on the accuracy of the engine used to generate, place, and rank the conformations into the target binding site [44]. A crucial step in the setup of a docking experiment is therefore the selection of a proper docking protocol; that is, the combination of search algorithm and scoring function that yields the best accuracy achievable. In the following, two different strategies to select/evaluate docking protocols are described, according to the amount of information available on the target of interest.

Evaluating Docking Performance through the Quality Descriptor-Driven Benchmark

When ligand-bound X-ray structures are available, a common practice is to evaluate the performance of different protocols in reproducing the experimental findings through a validation approach best known as a benchmark study. The accuracy of the protocols in benchmark studies is generally evaluated through the calculation of root mean square deviation (*RMSD*) values (Box 1) between the generated docking poses and the X-ray binding mode. Along with the classic *RMSD* estimate, we have recently proposed an alternative evaluation metric: the Protocol Score [45]. This *RMSD*-based descriptor also accounts for the average *RMSD* value returned by the docking protocols ($RMSD_{ave}$), the crystal structure resolution (R), and the number of conformations generated by each considered docking protocol with a *RMSD* value lower than R ($N^{(RMSD < R)}$). After docking simulations have been conducted, a 0–3 score is assigned to each docking protocol tested according to the criteria reported in Figure 3. The scores are converted into a color code and the data visualized as colored maps (Figure 3): white and light green spots correspond to poor protocol performances; dark green spots highlight

Box 1. Docking Poses Analysis Metrics

RMSD is a metric used to quantify to what extent the conformations of the generated docking poses differ from a reference structure (usually the X-ray observed binding mode). *RMSD* values are expressed in Ångströms and are computed by comparing the deviation of the coordinates of the heavy atoms with the following equation:

$$RMSD(\boldsymbol{a},\boldsymbol{b}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((a_{ix}-b_{ix})^2+(a_{iy}-b_{iy})^2+(a_{iz}-b_{iz})^2\right)} \quad [I]$$

where *a* and *b* are the sets of *n* heavy atoms of the docking pose and the reference structure, respectively (Figure IA).

IEF is an alternative metric to evaluate the quality of docking poses. From a mathematical viewpoint, this represents the contributions of each protein residue involved in the binding with the ligand to the IE. In particular, IEele (Figure IB) is computed on the basis of the non-bonded electrostatic interaction energy term of the force field and is expressed in kcal·mol^{-1}.

IEhyd (Figure IC) is computed in a less trivial way on the basis of a hydrophobic interaction term based on contact surfaces. The term takes into account for each ligand–protein atom pair the hydrophobicity character, the atom type (hybridization), and the reciprocal orientation of the two atoms. The outcome score is expressed in arbitrary units (the higher the better).

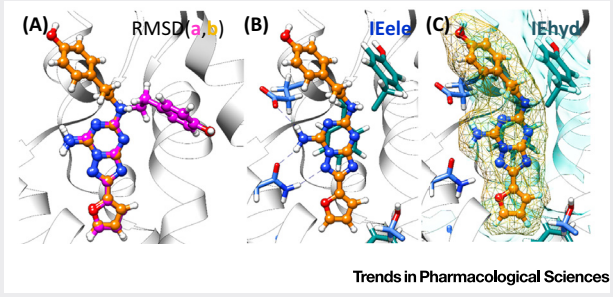


Figure I. Approaches to Evaluate Ligand Binding.

good protocols yielding good performances; and blue spots identify the best protocol tested. The Protocol Score metric has been used to evaluate 16 different docking protocols in a benchmark study focused on hA_{2A} AR [45] and has been recently embedded into DockBench 1.0, a platform to perform automated docking benchmarks [46].

Selection of docking protocols merely based on *RMSD* estimates, however, can lead to misleading results, as this metric carries no information about the correctness of the established ligand–protein interactions. Therefore, alternative metrics based on so-called interaction fingerprints (IFs) have been developed [39,40,47,48]. The underlying idea is to select docking poses that satisfy a previously defined binding hypothesis, such as the establishment of key interactions or a certain degree of surface complementarity with the binding pocket. IFs can be visualized as binary strings (Figure 4A) indicating the presence/absence of either automatically generated [48] or user-selected [47] ligand–protein interactions or as semiquantitative heat-like maps [e.g., interaction energy (IE) maps (IEMs)]. In the bit-string representation, the ligand–protein interactions are encoded in a 1D vector enabling fast comparison of large numbers of docking poses as well as the computation of similarity indexes. Such types of data are particularly suitable for VS purposes. IEMs (Figure 4B) are derived by computing IE fingerprints (IEFs) (Box 1) – that is, per-residue electrostatic and hydrophobic contributions to the IE (IEele and IEhyd, respectively) – for selected residues for both the crystallographic binding mode and the docking poses. IEMs are then generated by plotting the ligand identifiers against the key residues involved in the binding with a color code reflecting the IE values (the more intense the color, the stronger the interaction). The IEMs enable fast visual inspection and comparison of a limited number of docking poses and are more suitable for SAR rationalization [45].

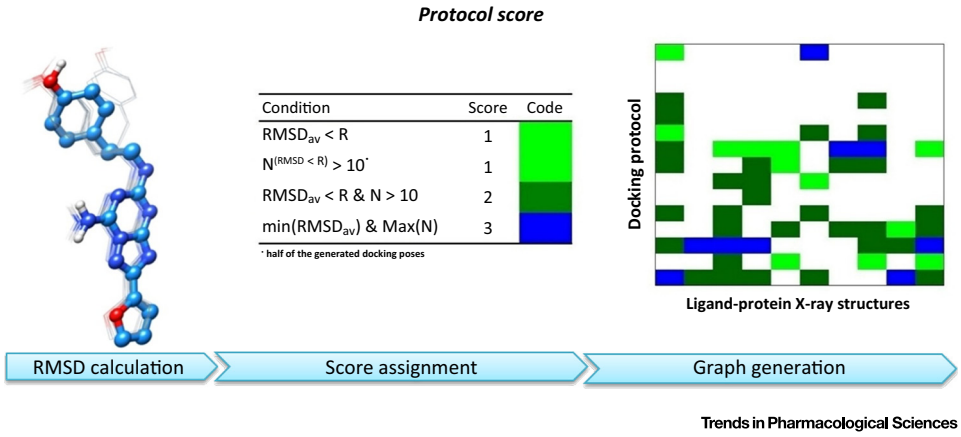


Figure 3. Schematic Representation of Workflows to Derive the Protocols Score. Adapted from [45].

Assessing the Quality of Homology Models through Blind Assessments

If X-ray structures are not available for comparison, the selection of a proper docking protocol is not a straightforward task. To this end, community-wide blind GPCR dock assessments offer a valuable opportunity to evaluate the status of method development in ligand–GPCR structure prediction. To date, three rounds of assessment have been conducted [49–51] that highlighted successful strategies as well as methodological shortcomings that need further improvements. Overall, the GPCR dock challenges revealed that: (i) reliable homology modeling requires at least 35–40% sequence identity between target and template; (ii) model selection/refinement is most successful when guided by available experimental data about the target and known ligands; and (iii) the use of automated procedures and sophisticated algorithms cannot disregard expert knowledge and chemical intuition. Focusing on the prediction of ligand–GPCR binding mode, the results indicated that accurate prediction of interactions with ECL regions remains challenging. In the orthosteric binding site, errors in binding mode predictions are mainly due to residue side chain packing arising from incorrect sequence alignments and/or unsuitable template selections during the homology modeling procedure. Regarding the accuracy of docking protocols, a common source of error is the neglect of water molecules directly or indirectly involved in ligand interactions [49]. These shortcomings were also underlined in a systematic analysis aimed at evaluating the performance of rigid and flexible receptor docking methodologies on both X-ray structures and homology models of several class A GPCRs [52].

Accounting for the Roles of Receptor Flexibility and Water Molecules in Ligand Binding

As discussed above, the conventional molecular docking approaches suffer from several limitations [53]. Although docking is a valuable method to obtain insights on the final stage of ligand–protein recognition, it lacks a description of three fundamental aspects that might play a significant role in ligand binding: a realistic microenvironment (particularly for membrane proteins), protein flexibility, and water molecule-mediated interactions. MD is a methodology that better accounts for these critical aspects and MD simulations (Figure 2C) are now a powerful tool widely used in medicinal chemistry and drug discovery [54]. Setting up MD simulations on GPCRs is a task that requires a certain degree of expertise. In particular, GPCR insertion into a phospholipid bilayer is not trivial, as it requires knowledge of protein spatial arrangement in the membrane. In the past years, several tools to aid users in predicting protein orientation in the bilayer as well as to embed the structures into membrane model systems have been made available to the scientific community, thus promoting the simulation of GPCRs in a realistic environment as a task within everyone's reach (Box 2).

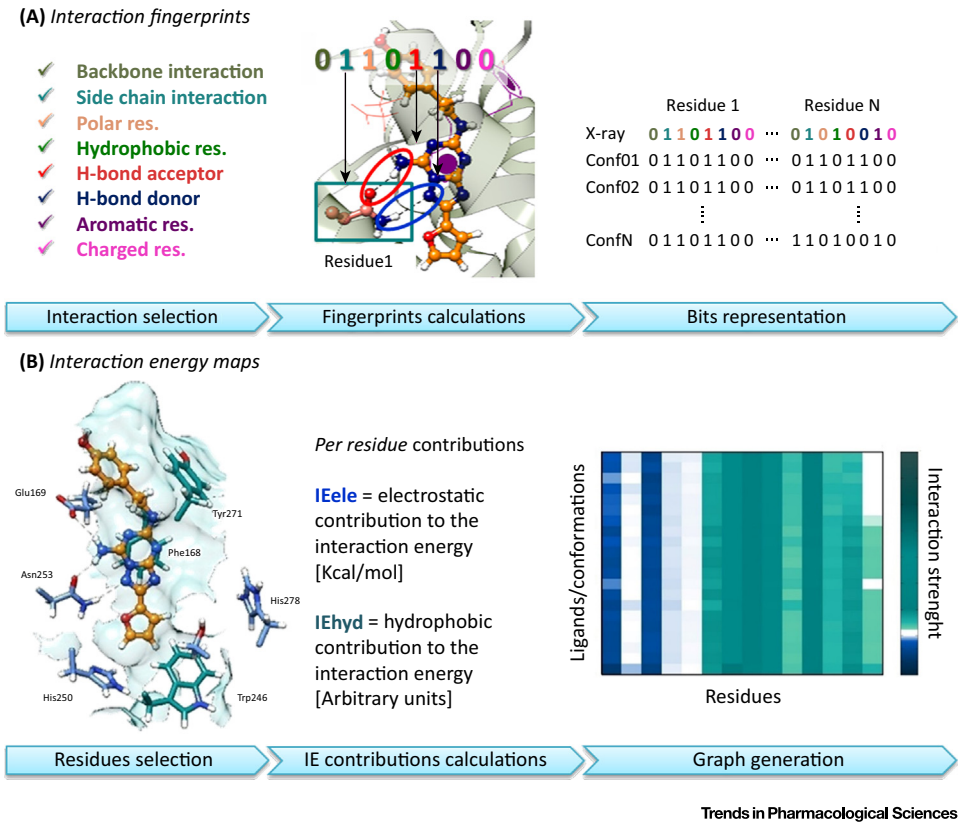


Figure 4. Schematic Representation of Workflows to Derive (A) Interaction Fingerprints and (B) Interaction Energy Maps. Per-residue electrostatic contribution (*IEele*) values: kcal $Å^{-1}$·mol^{-1}; per-residue hydrophobic contribution (*IEhyd*) values: arbitrary units. Adapted from [45].

An anticipated, to overcome the limitations related to the docking method, particularly regarding the contributions arising from water molecule-mediated interactions and protein side chain flexibility, several approaches based on membrane MD simulations have been recently introduced and applied to class A GPCRs.

Following the Dynamics of Ligand–Receptor Interactions

The underlying idea of the dynamic scoring function (DSF) (Figure 5) [55] is to provide, starting from a docking pose, a 'dynamic' estimate of ligand–protein interactions while accounting for the interplay of water molecules and protein side chain flexibility. Mathematically, the DSF is the cumulative sum of electrostatic and hydrophobic contributions to ligand–protein interactions corrected for ligand fluctuation (*RMSD*) with respect to the starting position ($wDSF_{ele}$ and $wDSF_{hyd}$, respectively). The values are computed for residues within a range of 4.5 Å from the ligand and are calculated as follows:

$$wDSF_{ele} = \frac{\Sigma_{t=0}^{n} IE_{ele}}{RMSD} \quad [1]$$

$$wDSF_{hyd} = \frac{\Sigma_{t=0}^{n} IE_{hyd}}{RMSD} \quad [2]$$

The resulting graphs (Figure 5) obtained by plotting *wDSF* against simulation time enable a graphical comparison of the relative stability of docking poses. This representation can help in detecting and validating the feasibility of alternative binding conformations proposed by the

Box 2. Setting Up Membrane MD Simulations of GPCRs

The first step in the setup of a membrane–protein system is the determination of the protein orientation into the bilayer through identification of the 'hydrophobic belt' (HB); that is, the protein surface exposed to the amphipathic portion of the membrane. This task can be performed through: (i) manual; (ii) semiautomatic; or (iii) automatic alignment. In the manual procedure, the user identifies the HB by visual inspection and manually rotates/translates the protein to orient it. In semiautomatic alignment – the basis of the OPM database [73] – an algorithm detects the HB and the protein is rotated/translated manually. In the automatic procedure, an algorithm automatically detects the HB and orients the protein. The CHARMM-GUI Membrane Builder server [74] and the LAMBADA [75] and MEMEBED [76] software packages, although differing in their underlying algorithms, are based on this approach. The oriented protein can be embedded into the membrane according to: (i) the replacement method; or (ii) the insertion method (Figure I). The replacement approach builds the membrane around the protein by placing dummy atoms at predicted lipid head groups' locations and replacing them with lipids. This strategy – the basis of the CHARMM-GUI membrane builder server [74] – allows fine control of system size but requires long equilibrations. The insertion approach places the protein into a pre-equilibrated bilayer leaflet by removing overlapping lipids. This strategy requires shorter equilibrations but limits the system size to the dimensions of available patches. In recent years several insertion methods have been developed. The GRIFFIN approach [77] generates a void in the leaflet and optimizes the system by applying repulsive forces mimicking the protein shape to the surrounding lipids. The g_membed tool [78] creates a small void in the leaflet and inserts into it a flattened protein gradually inflated back to the original size. The InflateGRO2 [75] procedure expands the lipid bilayer by bringing the system back to normal size through alternate compressions and minimizations. All of these methods are usually restricted to specific software/force fields. Recently, a universal strategy has also been proposed [79] comprising the application of a lateral pressure that pushes the protein into the membrane while proper constraints ensure its folding.

Regarding the homology models, web services with available pre-equilibrated membrane–protein systems are now emerging. The latest release of the MemProtMD database [80] – the largest and most updated repository – represents a great advance toward growing exploitation of membrane MD simulations by the scientific community.

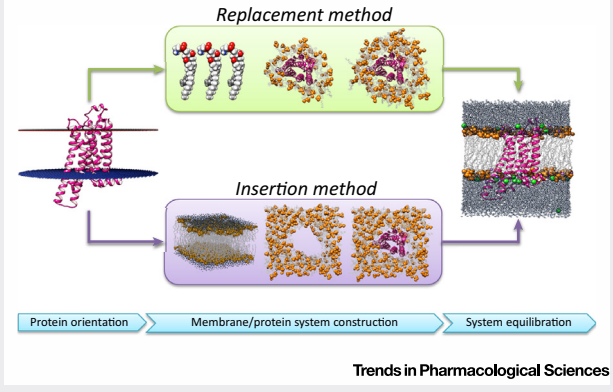
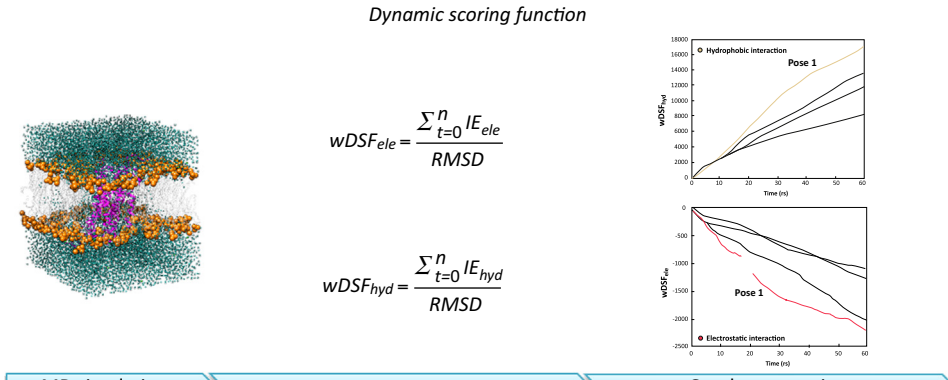


Figure I. Membrane Molecular Dynamics (MD) Simulations of G-Protein-Coupled Receptors (GPCRs).

docking algorithm. We have exploited this feature to support an apparently less plausible binding mode of a series of 5-alkylaminopyrazolo[4,3-*e*]1,2,4-triazolo[1,5-*c*]pyrimidine at the hA_3 AR [56] and tested the applicability of this approach by participating in the community-wide 2013 GPCR Dock assessment [51]. Several alternative $5HT_{2B}$/ergotamine complexes suggested by the docking protocol were submitted to membrane MD simulations. The best final poses were selected according to the outcomes of the DSF analysis. The overall performance of our approach can be appreciated by browsing the final dock assessment ranking list, which placed it eighth among 254, suggesting the portability of our approach to homology models as well as to other GPCRs.

Degree of Happiness of Water Molecules during Ligand Binding

The contribution of water molecules to protein–ligand binding can arise from at least two different events. Water molecules can mediate ligand–protein interaction through hydrogen bond networks [57] or can be displaced by the ligand from the protein surface and released to the bulk [57,58]. The protein regions where water molecules playing a key role in ligand binding mostly reside are defined as 'hot spots' and the ability to predict their location is of utmost importance in



Trends in Pharmacological Sciences

Figure 5. Schematic Representation of Workflows to Derive the Weighted Dynamic Scoring Function (*wDSF*). Adapted from [55].

drug discovery. In the GPCR research field, the importance of taking explicitly into account the perturbation of water networks resulting from ligand binding has been clearly demonstrated [59,60] and the modulation of internal water channels has recently emerged as an alternative approach to develop drug candidates [60]. Owing to the relevance of this contribution to ligand binding, in the past years several specialized tools based on MD simulations have been developed to aid compound synthesis prioritization in hit-to-lead or lead optimization programs. The WaterMap software [61,62] locates water molecules solvating the binding site by providing estimates of their energetic properties through short MD simulations in explicit solvent followed by a statistical thermodynamic analysis. 'Unhappy' water molecules are those with associated predicted positive free energy values with respect to water molecules in the bulk solvent. In an analysis focused on the binding of a series of triazine analogs at the A_{2A} AR, unhappy water molecules trapped between the ligand and the protein have been linked to a decrease of molecule residence time [59].

Another approach based on the WaterFLAP protocol [63] generates and scores water networks for apo and ligand–protein structures through the iterative use of OH2 GRID water hot spots. The initial networks are subjected to short (20 ps) MD simulations and water molecules in the proximity of the ligand are rescored in a final stage using water OH2 and CRY probes. In a compared analysis, agreement has been found for both the predicted locations and energetic estimates of unhappy water molecules obtained from the two different methodologies for the triazine series [63]. WaterFLAP analysis has also provided a rationalization of the 'magic-methyl' effect of antagonists in the chromone series at the A_{2A} AR [64].

An alternative MD-based method aimed at locating protein hot spots has also been recently proposed by us [65]. Our approach inspects the time-dependent variation of fluid dynamics properties of water molecules as a consequence of ligand binding, by creating a box surrounding the binding site (Figure 6) and dividing it into a 3D grid. During the MD simulations, the diffusion of water molecules in each grid cell is followed. The averaged positions of water molecules having *RMSF* values below 1.4 Å are projected into a bidimensional grid. The overlap of the grids yields a water fluid dynamics (WFD) map with cells colored according to the residence time of water molecules on a 0–100% scale. White zones (0%) are occupied by molecules with a residence time equivalent to bulk, whereas blue regions (100%) are occupied by molecules showing the maximum residence time of the considered trajectory. These maps help in graphically inspecting the water distribution inside the binding site and identifying protein

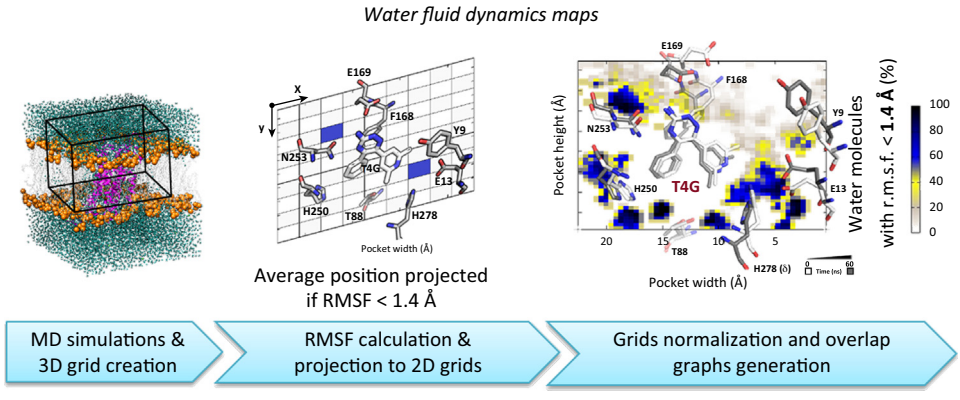


Figure 6. Schematic Representation of Workflows to Derive Water Fluid Dynamics (WFD) Maps. Adapted from [65].

regions where scaffold decorations avoid the disruption of key water molecules networks. We also investigated with this approach the binding of the triazine series at the hA_{2A} AR [65]. Our results were in agreement with the experimental data and with the analyses reported above. WFD maps suggested that the hydrophobic moieties of ligands with a short residence time unhinge a cluster of water molecules whose location agrees with the placement of unhappy water molecules predicted by the WaterMap and WaterFLAP protocols [63,64]. The analysis of WFD maps of the ligand exhibiting slower off-rate kinetics highlights the presence of three stable clusters of water molecules, suggesting that the increase in binding affinity might arise from the network of water molecules around the ligand [65].

Exploring Ligand–Receptor Recognition

Knowledge of the ligand–GPCR recognition pathway would facilitate the development of drug candidates with better pharmacodynamic profiles. Unfortunately, simulating this process remains a challenging task because it requires classical MD experiments over a long, micro-second timescale [26,66,67] that is affordable only with specialized supercomputers [68]. Recently it was demonstrated that a broader community might achieve long-timescale MD simulations by exploiting cloud computing coupled with Markov state models, an approach based on the estimate of interconversion rates of protein conformational states extrapolated by multiple short, independent trajectories. This approach proved fruitful in providing insights on the activation pathways of the β_2 AR, a class A GPCR [28].

The probability of capturing ligand–GPCR binding or unbinding on an accessible timescale can be enhanced through the introduction of biased potentials that facilitate the crossing of energy barriers or the application of external forces on the ligand, respectively [69]. An alternative strategy that does not require the introduction of biases or external forces and enables exploration of the ligand–GPCR approach path on a nanosecond simulation timescale has been recently proposed by us [70]. The supervised MD (SuMD) approach exploits a tabu-like algorithm to monitor the distance between the center of mass of the ligand atoms and the receptor-binding site in short (600 ps) standard MD simulations (Figure 7, left panel). According to this strategy, an arbitrary number of distance points is collected 'on the flight' at regular intervals and fitted into a linear function $f(x) = mx$. If the slope (m) is negative, the ligand–receptor distance is likely to be shortened and the simulation is restarted from the last set of coordinates. Otherwise, the simulation is restored from the original set of coordinates and started over. The supervision is repeated until the ligand–receptor distance is less than 5 Å. The results of a SuMD

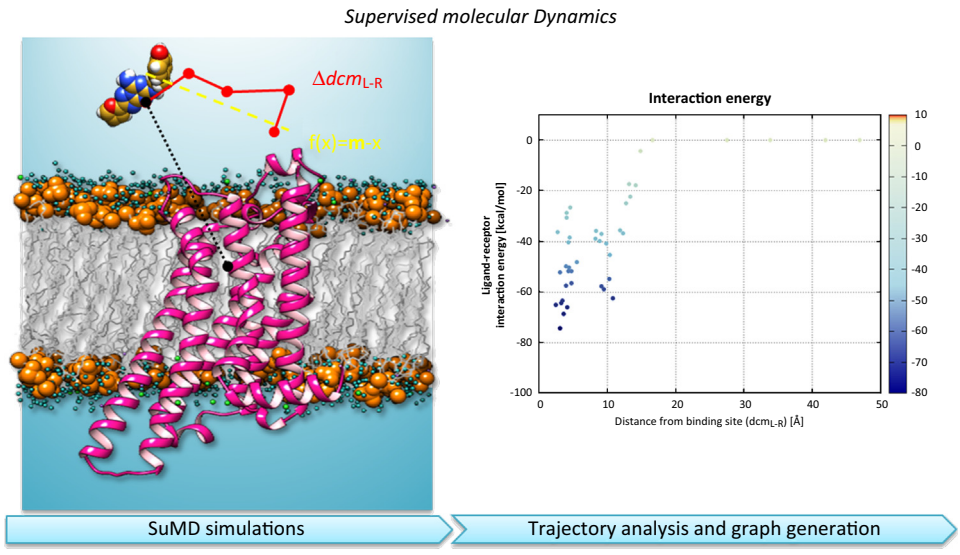


Figure 7. Schematic Representation of Supervised Molecular Dynamics (SuMD) Algorithm (Left) and the Outcome Ligand–Receptor Interaction Energy Landscape (right). Interaction energy values: kcal·mol^{-1}. Adapted from [70].

simulation are displayed in a graph reporting the IE toward the distance between the ligand and the binding site (Figure 7, right panel). We have recently applied the SuMD approach to interpret at the molecular level: (i) the binding of the natural agonist adenosine at the hA_2 AR by detecting and characterizing a possible energetically stable meta-binding site [71]; and (ii) the positive allosteric modulation mediated by LUF6000 toward the hA_3 AR by suggesting at least two possible mechanisms to explain the available experimental data [72].

Concluding Remarks

We have surveyed the recent advances in molecular modeling techniques aimed at elucidating the ligand–GPCR recognition process. The crystallographic revolution of the past decade and the advent of GPUs in the molecular modeling field have allowed the tuning of several new tools to assist the drug design procedure. The recently developed approaches enrich the pool of molecular modeling techniques currently available to disclose the factors influencing the ligand–GPCR recognition process and exploit three computational methodologies extensively used by modelers: homology modeling, molecular docking, and membrane MD simulations.

Most of the tools here presented hold the potential for future widespread use within the scientific community. Freely available web resources, full automation, and the acceleration of multistep procedures as well as the representation of results with easy-to-interpret graphs represent the key actions to encourage non-expert users to employ these new tools. Moreover, the possibility to rapidly reanalyze previously obtained outputs with some of the new approaches is further aimed at stimulating the interest of the scientific community.

Hopefully, through this review we have demonstrated the potential of these techniques in solving real-life problems in drug discovery and encouraged novice modelers to exploit them while embarking on the fascinating yet complicated task of ligand–GPCR recognition simulations. In the future, we foresee the extension and improvement of the applicability of these computational tools to address other fascinating open questions in the GPCR field (see Outstanding Questions) and we hope that these approaches, carefully integrated with all of the other experimental

Outstanding Questions

By which molecular mechanism do orthosteric and allosteric binders control GPCR functionality?

Are the kinetic parameters of ligand–receptor recognition, such as k_{on} and k_{off}, computationally addressable?

What is the real role of water in ligand binding and receptor activation?

What are the implications of phosphorylation and glycosylation in ligand binding and receptor activation?

What is the physiopathological meaning of monomer–oligomer (homo and/or hetero) receptor equilibrium?

Are there other second messengers involved in G-protein-alternative signaling pathways affecting ligand recognition?

GPCRs competencies, will broaden our perspectives in several scientific areas from molecular pharmacology to drug discovery.

Author Contributions

This review was written with contributions from all authors. All authors gave approval for the final version.

Acknowledgments

The work was conducted with financial support from the University of Padova, Italy and the Italian Ministry for University and Research, Rome, Italy. S.M. participates in the European COST Action CM1207 (GLISTEN).

References

1. Alexander, R.W. *et al.* (1975) Identification of cardiac β-adrenergic receptors by (minus) [3H]alprenolol binding. *Proc. Natl. Acad. Sci. U.S.A.* 72, 1564–1568
2. Dixon, R.A.F. *et al.* (1986) Cloning of the gene and cDNA for mammalian β-adrenergic receptor and homology with rhodopsin. *Nature* 321, 75–79
3. Cherezov, V. *et al.* (2007) High-resolution crystal structure of an engineered human 2-adrenergic G protein-coupled receptor. *Science* 318, 1258–1265
4. Pierce, K.L. *et al.* (2002) Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* 3, 639–650
5. Kristiansen, K. (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.* 103, 21–80
6. Shukla, A.K. *et al.* (2014) Emerging structural insights into biased GPCR signaling. *Trends Biochem. Sci.* 39, 594–602
7. Moreno, J.L. *et al.* (2013) G protein-coupled receptor heterocomplexes in neuropsychiatric disorders. *Prog. Mol. Biol. Transl. Sci.* 117, 187–205
8. O'Hayre, M. *et al.* (2014) Novel insights into G protein and G protein-coupled receptor signaling in cancer. *Curr. Opin. Cell Biol.* 27, 126–135
9. Sodhi, A. *et al.* (2004) Viral hijacking of G-protein-coupled-receptor signalling networks. *Nat. Rev. Mol. Cell Biol.* 5, 998–1012
10. Conn, P.J. *et al.* (2014) Opportunities and challenges in the discovery of allosteric modulators of GPCRs for treating CNS disorders. *Nat. Rev. Drug Discov.* 13, 692–708
11. Lundstrom, K. (2006) Latest development in drug discovery on G protein-coupled receptors. *Curr. Protein Pept. Sci.* 7, 465–470
12. Fredriksson, R. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63, 1256–1272
13. Milić, D. and Veprintsev, D.B. (2015) Large-scale production and protein engineering of G protein-coupled receptors for structural studies. *Front. Pharmacol.* 6, 66
14. Stevens, R.C. *et al.* (2012) The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nat. Rev. Drug Discov.* 12, 25–34
15. Chun, E. *et al.* (2012) Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* 20, 967–976
16. Hino, T. *et al.* (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* 482, 237–240
17. Warne, T. *et al.* (2008) Structure of a β1-adrenergic G-protein-coupled receptor. *Nature* 454, 486–491
18. Shonberg, J. *et al.* (2015) GPCR crystal structures: medicinal chemistry in the pocket. *Bioorg. Med. Chem.* 23, 3880–3906
19. Jacobson, K.A. and Costanzi, S. (2012) New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. *Mol. Pharmacol.* 82, 361–371
20. Tautermann, C.S. (2014) GPCR structures in drug design, emerging opportunities with new structures. *Bioorg. Med. Chem. Lett.* 24, 4073–4079
21. Salon, J.A. *et al.* (2011) The significance of G protein-coupled receptor crystallography for drug discovery. *Pharmacol. Rev.* 63, 901–937
22. Guerrero, G.D. *et al.* (2014) A performance/cost evaluation for a GPU-based drug discovery application on volunteer computing. *Biomed. Res. Int.* 2014, 1–8
23. Buch, I. *et al.* (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. *J. Chem. Inf. Model.* 50, 397–403
24. Anderson, J.A. *et al.* (2008) General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* 227, 5342–5359
25. Mashimo, T. *et al.* (2013) Molecular dynamics simulations accelerated by GPU for biological macromolecules with a non-Ewald scheme for electrostatic interactions. *J. Chem. Theory Comput.* 9, 5599–5609
26. Dror, R.O. *et al.* (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13118–13123
27. Selvam, B. *et al.* (2012) Comparison of dynamics of extracellular accesses to the β1 and β2 adrenoceptors binding sites uncovers the potential of kinetic basis of antagonist selectivity: selectivity at the adrenoceptors. *Chem. Biol. Drug Des.* 80, 215–226
28. Kohlhoff, K.J. *et al.* (2013) Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* 6, 15–21
29. Paoletta, S. *et al.* (2010) Receptor-driven identification of novel human A_3 adenosine receptor antagonists as potential therapeutic agents. *Methods Enzymol.* 485, 225–244
30. Katritch, V. *et al.* (2012) Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* 33, 17–27
31. Simms, J. *et al.* (2009) Homology modeling of GPCRs. In *G Protein-Coupled Receptors in Drug Discovery* (Vol. 552) (Leifert, W.R., ed.), pp. 97–113, Humana Press
32. Horn, F. (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* 31, 294–297
33. Kowalsman, N. and Niv, M.Y. (2014) GPCR & Company: databases and servers for GPCRs and interacting partners. In *G Protein-Coupled Receptors – Modeling and Simulation* (Vol. 796) (Filizola, M., ed.), pp. 185–204, Springer
34. Rodríguez, D. *et al.* (2012) Molecular modelling of G protein-coupled receptors through the web. *Mol. Inform.* 31, 334–341
35. Floris, M. *et al.* (2012) Adenosiland: walking through adenosine receptors landscape. *Eur. J. Med. Chem.* 58, 248–257
36. Isberg, V. *et al.* (2014) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 42, D422–D425
37. Floris, M. *et al.* (2013) Implementing the "Best Template Searching" tool into Adenosiland platform. *Silico Pharmacol.* 1, 25
38. Gutiérrez-de-Terán, H. *et al.* (2013) Characterization of the dynamic events of GPCRs by automated computational simulations. *Biochem. Soc. Trans.* 41, 205–212
39. Kelly, M.D. and Mancera, R.L. (2004) Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design. *J. Chem. Inf. Model.* 44, 1942–1951

40. Kroemer, R.T. *et al.* (2004) Assessment of docking poses: interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J. Chem. Inf. Model.* 44, 871–881
41. Marcou, G. and Rognan, D. (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* 47, 195–207
42. Kooistra, A.J. *et al.* (2013) From heptahelical bundle to hits from the haystack. *Methods Enzymol.* 522, 279–336
43. Kooistra, A.J. *et al.* (2014) From three-dimensional GPCR structure to rational ligand discovery. In *G Protein-Coupled Receptors – Modeling and Simulation* (Vol. 796) (Filizola, M., ed.), pp. 129–157, Springer
44. Houston, D.R. and Walkinshaw, M.D. (2013) Consensus docking: improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.* 53, 384–390
45. Ciancetta, A. *et al.* (2014) Alternative quality assessment strategy to compare performances of GPCR–ligand docking protocols: the human adenosine A_{2A} receptor as a case study. *J. Chem. Inf. Model.* 54, 2243–2254
46. Cuzzolin, A. *et al.* (2015) DockBench: an integrated informatic platform bridging the gap between the robust validation of docking protocols and virtual screening simulations. *Molecules* 20, 9977–9993
47. Deng, Z. *et al.* (2004) Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein–ligand binding interactions. *J. Med. Chem.* 47, 337–344
48. Da, C. and Kireev, D. (2014) Structural protein–ligand interaction fingerprints (SPLIF) for structure-based virtual screening: method and benchmark study. *J. Chem. Inf. Model.* 54, 2555–2561
49. Michino, M. *et al.* (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* 8, 455–463
50. Kufareva, I. *et al.* (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* 19, 1108–1126
51. Kufareva, I. *et al.* (2014) Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* 22, 1120–1139
52. Beuming, T. and Sherman, W. (2012) Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. *J. Chem. Inf. Model.* 52, 3263–3277
53. Warren, G.L. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J. Med. Chem.* 49, 5912–5931
54. Durrant, J.D. and McCammon, J.A. (2011) Molecular dynamics simulations and drug discovery. *BMC Biol.* 9, 71
55. Sabbadin, D. *et al.* (2014) Bridging molecular docking to membrane molecular dynamics to investigate GPCR–ligand recognition: the human A_{2A} adenosine receptor as a key study. *J. Chem. Inf. Model.* 54, 169–183
56. Federico, S. *et al.* (2012) Exploring the directionality of 5-substitutions in a new series of 5-alkylaminopyrazolo[4,3-*e*]1,2,4-triazolo[1,5-*c*]pyrimidine as a strategy to design novel human A_3 adenosine receptor antagonists. *J. Med. Chem.* 55, 9654–9668
57. Lu, Y. *et al.* (2007) Analysis of ligand-bound water molecules in high-resolution crystal structures of protein–ligand complexes. *J. Chem. Inf. Model.* 47, 668–675
58. Snyder, P.W. *et al.* (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17889–17894
59. Bortolato, A. *et al.* (2013) Water network perturbation in ligand binding: adenosine A_{2A} antagonists as a case study. *J. Chem. Inf. Model.* 53, 1700–1713
60. Yuan, S. *et al.* (2014) Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat. Commun.* 5, 4733
61. Beuming, T. *et al.* (2012) Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins* 80, 871–883
62. Wang, L. *et al.* (2011) Ligand binding to protein-binding pockets with wet and dry regions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1326–1330
63. Mason, J.S. *et al.* (2013) High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks. *Silico Pharmacol.* 1, 23
64. Andrews, S.P. *et al.* (2014) Structure-based drug design of chromone antagonists of the adenosine A_{2A} receptor. *Med. Chem. Commun.* 5, 571
65. Sabbadin, D. *et al.* (2014) Perturbation of fluid dynamics properties of water molecules during G protein-coupled receptor–ligand recognition: the human A_{2A} adenosine receptor as a key study. *J. Chem. Inf. Model.* 54, 2846–2855
66. Buch, I. *et al.* (2011) Complete reconstruction of an enzyme–inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10184–10189
67. Dror, R.O. *et al.* (2013) Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* 503, 295–299
68. Shaw, D.E. *et al.* (2009) Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pp. 1 Association for Computing Machinery
69. Johnston, J.M. and Filizola, M. (2014) Beyond standard molecular dynamics: investigating the molecular mechanisms of G protein-coupled receptors with enhanced molecular dynamics methods. In *G Protein-Coupled Receptors – Modeling and Simulation* (Vol. 796) (Filizola, M., ed.), pp. 95–125, Springer
70. Sabbadin, D. and Moro, S. (2014) Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR–ligand recognition pathway in a nanosecond time scale. *J. Chem. Inf. Model.* 54, 372–376
71. Sabbadin, D. *et al.* (2015) Exploring the recognition pathway at the human A_{2A} adenosine receptor of the endogenous agonist adenosine using supervised molecular dynamics simulations. *Med. Chem. Commun.* Published online April 20, 2015. http://dx.doi.org/10.1039/C5MD00016E
72. Deganutti, G. *et al.* (2015) Understanding allosteric interactions in G protein-coupled receptors using supervised molecular dynamics: a prototype study analysing the human A3 adenosine receptor positive allosteric modulator LUF6000. *Bioorg. Med. Chem.* 23, 4065–4071
73. Lomize, M.A. *et al.* (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22, 623–625
74. Wu, E.L. *et al.* (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* 35, 1997–2004
75. Schmidt, T.H. and Kandt, C. (2012) LAMBADA and InflateGRO2: efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations. *J. Chem. Inf. Model.* 52, 2657–2669
76. Nugent, T. and Jones, D.T. (2013) Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics* 14, 276
77. Staritzbichler, R. *et al.* (2011) GRIFFIN: a versatile methodology for optimization of protein–lipid interfaces for membrane protein simulations. *J. Chem. Theory Comput.* 7, 1167–1176
78. Wolf, M.G. *et al.* (2010) g_membed: efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. *J. Comput. Chem.* 31, 2169–2174
79. Javanainen, M. (2014) Universal method for embedding proteins into complex lipid bilayers for molecular dynamics simulations. *J. Chem. Theory Comput.* 10, 2577–2582
80. Stansfeld, P.J. *et al.* (2015) MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure* 23, 1350–1361