CHAPTER

16 Computational studies of G protein-coupled receptor complexes: Structure and dynamics

Ozge Sensoy*, Jose G. Almeida†, Javeria Shabbir*, Irina S. Moreira†,‡, Giulia Morra§,¶,1

*Istanbul Medipol University, The School of Engineering and Natural Sciences, Istanbul, Turkey
†CNC—Center for Neuroscience and Cell Biology, Universidade de Coimbra, Coimbra, Portugal
‡Bijvoet Center for Biomolecular Research, Faculty of Science—Chemistry, Utrecht University, Utrecht, The Netherlands
§Weill-Cornell Medical College, Cornell University, New York, New York, United States
¶ICRM-CNR Istituto di Chimica del Riconoscimento Molecolare, Consiglio Nazionale delle Ricerche, Milano, Italy
1Corresponding author: e-mail address: giulia.morra@icrm.cnr.it

CHAPTER OUTLINE

Methods in Cell Biology, Volume 142, ISSN 0091-679X, https://doi.org/10.1016/bs.mcb.2017.07.011

Abstract

G protein-coupled receptors (GPCRs) are ubiquitously expressed transmembrane proteins associated with a wide range of diseases such as Alzheimer's, Parkinson, schizophrenia, and also implicated in in several abnormal heart conditions. As such, this family of receptors is regarded as excellent drug targets. However, due to the high number of intracellular signaling partners, these receptors have a complex interaction networks and it becomes challenging to modulate their function.

Experimentally determined structures give detailed information on the salient structural properties of these signaling complexes but they are far away from providing mechanistic insights into the underlying process. This chapter presents some of the computational tools, namely molecular dynamics, molecular docking, and molecular modeling and related analyses methods that have been used to complement experimental findings.

1 INTRODUCTION

G protein-coupled receptors (GPCRs) are involved in many pathophysiological pathways of crucial diseases such as cancer, Alzheimer, Parkinson's disease, obesity, etc., and thus are target of approximately 40% of all currently prescribed medicinal drugs. Either restoring the function of the receptor in disease-associated pathways or biasing the receptor toward a specific signaling pathway requires knowledge of structural and dynamic properties of the system at the molecular level. However, the presence of a high number of intracellular partners (Hermans, 2003), giving rise to intricate interaction networks, make any pharmacological approach challenging.

The physiological function of GPCRs is mediated by protein–protein interactions formed between the receptor and their effectors, such as heterotrimeric G-proteins or Arrestin (Arr-s) proteins (Han, Moreira, Urizar, Weinstein, & Javitch, 2009; Moreira et al., 2010). One of the most important concepts that have emerged from recent findings is "functional selectivity," or the correlation between binding of a specific ligand to the receptor and activation of a specific signaling pathway, mediated by either G-protein or Arr-s (Fig. 1). The structural mechanism of functional selectivity has not been elucidated yet. Key developments in the quest for understanding the coupling between the receptor and its effector were the crystal structures of the β2-adrenergic receptor (β2AR) in complex with the G protein (Gs) (PDBid: 3SN6; Rasmussen et al., 2011) and of Arrestin-1 coupled to opsin (PDBid: 4ZWJ; Kang et al., 2015). These structures revealed remarkable changes on the structural properties of the "active form" of the GPCR than those complexes obtained in the presence of only agonist. However, despite the insight into the conformational changes adopted by the receptor and provided by such complexes, the underlying molecular mechanism of the functional selection process remains largely elusive. Moreover, the determination of structures via experimental methods is not straightforward, due to some inherent problems regarding the size of the complexes studied (e.g., by nuclear magnetic resonance), the accurate reproduction of the membrane environment, acquisition of the data itself (e.g., X-ray crystallography), and the GPCR expression levels (a problem observed across most experimental structure determination techniques) (Carpenter, Beis, Cameron, & Iwata, 2008; Ghosh, Kumari, Jaiman, & Shukla, 2015). In this respect, computational techniques have become indispensable tools to complement experimental findings. Modeling studies of relevant complexes of various GPCRs with G protein and Arrestin can be guided by the existing high-resolution information and illuminate the interface properties. Another powerful computational technique, molecular dynamics (MD), offers a useful approach to complement structural information and gives mechanistic insights that are not provided by static structures, and also helps make qualitative and semiquantitative predictions. They are particularly suited for addressing questions on GPCR activation, ligand-induced conformational changes, perturbations induced by mutants, and allosteric modulation. In this chapter, we will therefore focus on: (i) the prediction of protein–protein interaction interfaces and on (ii) MD simulations as

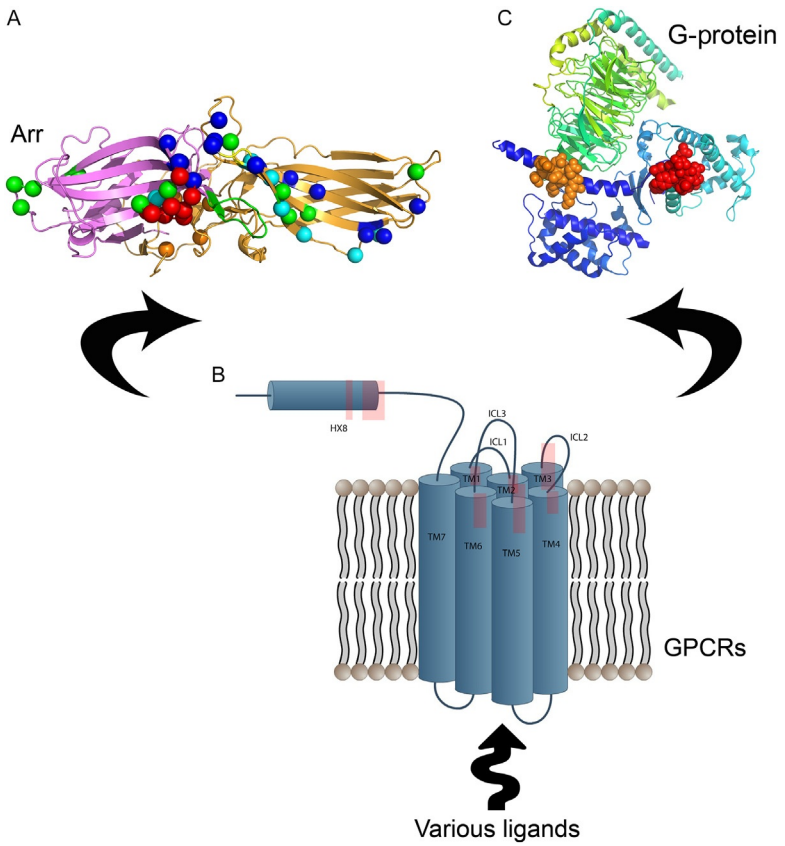

FIG. 1

Structural representation of: (A) Arr. The N- and C-domains, the "finger loop," the "neighboring loop," and the "TYR loop" are colored in *purple*, *orange*, *red*, *green*, and *yellow*, respectively. We have highlighted the Cα atom of the polar core residues (*black*), as well as residues known to be involved in GPCR binding. In *cyan* are the interfacial residues common to Arr2 and Arr3 whereas *blues* are the ones which are important for Arr2 and for Arr1/rhodopsin coupling. (B) G protein-coupled receptor. Key regions for GPCR/G protein and Arr coupling are represented by a *red square* (*different sized squares* are related to the dimension of the determinant region). (C) G protein. The 10 amino acid residues important for coupling of G protein to the receptor are shown in *red* and in vdW representation. The binding motif at the N-terminus is shown *orange*.

examples of two widely used powerful tools for providing high-resolution structural and mechanistic models for GPCR–effectors interactions. For both methods, we will illustrate the main theoretical background, the questions that can be addressed, the data production protocol, and the main analysis methods, in the context of modeling the function of GPCRs and their partners.

2 THEORY

2.1 CONSTRUCTION AND ANALYSIS OF INTERFACES IN GPCR/EFFECTOR COMPLEXES

In this section, we present a computational "metamethod" that makes use of already developed techniques for the characterization of GPCR–effector protein structures. In order to fully characterize the GPCR–partner interaction, the starting point is a 3D structure at atomic resolution of the complex. In the absence of experimentally solved conformations, the 3D structures of the constituents of the complex are separately determined through homology modeling, which is followed by the state-of-the-art protein–protein docking tools to achieve the interaction interface of the complex. As an example, a GitHub repository has been created, which includes a protocol to construct possible models of dopamine 1 receptor (D_1R) complex with several different G proteins (available at: https://github.com/IrinaMoreira/gpcr-comparative-analysis).

2.1.1 Homology modeling of a GPCR

Homology modeling is a procedure that generates a previously unknown protein structure by "fitting" its sequence (target) into a known structure (template), given a certain level of sequence homology (at least 30%) between target and template. First, the sequences of the template structure(s) should be retrieved using multiple alignment. Several multiple-sequence alignment (MSA) software applications and webservers, namely MUSCLE (Edgar, 2004), Clustal Omega (Sievers & Higgins, 2014), BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990; Camacho et al., 2009), PSI-Search (Li et al., 2012), and FASTA (Pearson, 2014), can be used for this task. After finding the sequences with high homology to the query model, the ones with available 3D structures must be filtered. Some available methods/webservers (e.g., MODELLER (Eswar et al., 2006; Webb & Sali, 2014), SWISS-MODEL (Biasini et al., 2014), FoldX (Schymkowitz et al., 2005), HHpred (Soding, Biegert, & Lupas, 2005), PRIME (Jacobson et al., 2004), and ROBETTA (Kim, Chivian, & Baker, 2004)), automatically search for the structural database, yielding templates with resolved structures, and their respective Protein Data Bank (PDB) ids (PDBids) (Berman et al., 2000). The corresponding coordinates of the template GPCR can be downloaded directly from the PDB (http://www.rcsb.org). There are a few online databases that provide specific template suggestions and homology modeling of the helical regions of GPCRs, which can be quite useful as an initial guess. Among them are GPCR-Sequence-Structure Feature_Extractor (Worth, Kleinau, & Krause, 2009; Worth, Kreuchwig, Kleinau, & Krause, 2011) (SSFE) and GPCR-ModSim (Esguerra, Siretskiy, Bello, Sallander, & Gutierrez-de-Teran, 2016). Here, it is crucial to decide whether a single template or multiple templates will be used for homology modeling. To address this issue, the researcher should pay close attention to the structural similarity between the target and the template(s), the presence/absence of mutations as well as specific motifs. Out of the aforementioned

modeling tools, MODELLER (Eswar et al., 2006; Webb & Sali, 2014) is one of the most widely used software and it can be accessed both as a webserver (ModWeb, available at: https://modbase.compbio.ucsf.edu/modweb/) and as a Python library, which provides a customizable way of performing homology modeling.

There are also additional considerations, which should be taken into account when modeling a GPCR: (i) to perform the sequence alignment of template and target, paying particular attention to the TMs as these are the structural elements showing the highest level of conservation; (ii) to give special attention to the intracellular loop 3 (ICL3), which links TM5 and TM6, one of the most important regions for the selectivity of GPCR–G protein interactions (Kobilka & Schertler, 2008) but usually not solved or substituted by a fused lysozyme in a large number of the available GPCR crystals (Moreira, 2014); and (iii) to take into account the ECL2, which connects TM4 and TM5 and can adopt a variety of conformations depending on the activation state. In general, the extracellular and intracellular loops are characterized by having low sequence similarity and high variability in terms of the length, as well as dynamic heterogeneity. The alignment should be visually inspected to ensure that the most conserved residue of each TM, X.50 is well aligned (according to the Ballesteros & Weinstein numbering, Ballesteros & Weinstein, 1995). A detailed list of additional considerations regarding the modeling of GPCRs can be found at Costanzi (2012) and Esguerra et al. (2016).

When modeling an "active" GPCR, more than one experimentally resolved structures can be used as template, in particular the X-ray structure of GPCR complex with either G protein (PDBid 3SN6) (Rasmussen et al., 2011) or visual Arrestin (PDBid 4ZWJ) (Kang et al., 2015), the complex between the adenosine A_2R receptor and an engineered G protein (PDBid: 5G53) (Carpenter, Nehme, Warne, Leslie, & Tate, 2016), and the complex formed between a nanobody and the β2-adrenergic receptor (PDBid: 3P0G) (Rasmussen et al., 2011). This protocol was applied to modeling D_1R binding to the members of the G protein family (Gi1, Gi2, Go, Gslo, and Gssh) and available at: https://github.com/IrinaMoreira/gpcr-comparative-analysis.

2.1.2 Docking refinement

After predicting the 3D structures of each monomer, one aims at modeling the interface between GPCR and its partner. The easiest way to construct a 3D structure of the complex is to use High Ambiguity Driven biomolecular Docking (HADDOCK) (de Vries, van Dijk, & Bonvin, 2010; Dominguez, Boelens, & Bonvin, 2003; Van Zundert et al., 2016), which harbors a webserver including several different methods for protein–protein docking. After registration, one can freely use the refinement step of the full HADDOCK docking procedure, which would present the most realistic mode for the interface of the modeled GPCR–effector complex. HADDOCK can make use of information on the known interfacial residues. If none of those residues are available in the literature, CPORT (de Vries & Bonvin, 2011) can be used to predict interfacial residues. HADDOCK can also be utilized as standalone software, which enables a customizable approach for GPCR–effector docking. In Section 3, an example of the refinement step is given for the D_1R–G_s protein complex.

2.1.3 Protein–protein interface characterization

Chemical, biological, and physical properties governing the formation of specific GPCR complexes can be determined by means of various computational methods that are commonly used to investigate putative interfaces. In particular, they are H-bonds (HB), salt bridges (SB) (Xu, Harrison, & Eck, 1997), accessible surface area (ASA) (Miller, Janin, Lesk, & Chothia, 1987; Miller, Lesk, Janin, & Chothia, 1987), normal modes through normal mode analysis (NMA) (Niv & Filizola, 2008), and evolutionary conservation (EC) (Guharoy & Chakrabarti, 2005; Hu, Ma, Wolfson, & Nussinov, 2000). This section contains a list of approaches, webservers, and programs that can be used to analyze these structural and evolutionary features. All the information associated with individual residues (HB, SB, ASA, and EC) should be reported accordingly in the corresponding position of the aligned sequences in the format of a table—named aligned table (AT). Residues belonging to specific regions, e.g., transmembrane helices, loops, or interfacial residues, should be indicated in the table as well, through the use of a color code, for instance. An example of the color-coded assignment can be found in *1_HOMOLOGY MODELLING/SEQUENCE ALIGNMENT/DXR-Colors.xlsx*.

2.1.3.1 Structural features: Visual MD and CoCoMaps

Visual Molecular Dynamics (VMD) (Humphrey, Dalke, & Schulten, 1996, available at http://www.ks.uiuc.edu/Research/vmd) is a visualization, modeling, and analysis software, which provides important details on the 3D structure of proteins and will be mentioned many times during this chapter. For instance, SB can be calculated by means of one of the extensions (*Salt Bridges*), implemented within the software. The program identifies this structural property across the entire complex by considering distances within a cutoff, usually less than 3.5 Å, between polar atom pairs such as oxygen and nitrogen. When dealing with a homology model, it is important to consider slightly larger cutoffs to account for the intrinsic uncertainty of the model.

CoCoMaps (Vangone, Spinelli, Scarano, Cavallo, & Oliva, 2011) is a webserver (available at: https://www.molnac.unisa.it/BioTools/cocomaps/), which gives insights into four types of features: (i) ASA values and differences for interface residues, (ii) HB for interface residues, (iii) a global overview of the interface in terms of hydrophobic and hydrophilic interactions, and (iv) overall ASA values for polar and nonpolar residues. Values from (i)—surface area values and their differences in complex and monomer—are important in defining important regions of the interface of the complex (Moreira, 2015). The output for (i) in CoCoMaps comprises ASA values for both bound and unbound forms of the monomers, as well as the difference between the two values, which provides valuable insights on the interfaces of complexes. Data from (ii) can be summarized in two groups: (1) the actual distance of the HB, which is considered for analysis purposes and (2) distance between Cα of residues participating in those HB. Values for (iii) offer an easy way for comparing the interfaces of different GPCR–effector complexes in terms of hydrophobic–hydrophobic, hydrophobic–hydrophilic, and hydrophilic–hydrophilic interactions, while providing valuable insight on the total polar and nonpolar buried (with low ASA) area.

2.1.3.2 Evolutionary features: EVFold and Consurf

An important step toward the understanding of the structural basis of the coupling between a GPCR and its bound effector is to identify the level of EC of sequences at the interface (i.e., how replacements observed in the GPCR itself have influenced the replacements observed in the binding partner) in order to explore coevolving sites. The rationale behind this is that, whenever an interaction between two proteins is essential for a physiological function, any replacement at this site will limit possible replacements at the spatially interacting counterparts. Coevolution studies have been widely documented at the intramolecular level (Benner & Gerloff, 1991; Hatrick & Taylor, 1994; Zvelebil, Barton, Taylor, & Sternberg, 1987) and to a lower extent in the intermolecular context (Guharoy & Chakrabarti, 2005; Hu et al., 2000), offering an interesting approach in the field of protein–protein interaction (Choi, Yang, Choi, Ryu, & Kim, 2009). Coevolution should as such be regarded as not only a characterizing feature, but also as a "filter" for weighting other interactions (HB and ASA). By considering only evolutionarily conserved residues, we can determine both the common residues across several different interfaces and those conserved and unique to specific interfaces, thereby mediating the formation of different GPCR–effector complexes (see Section 3.1.6).

EVFold (Hayat, Sander, Marks, & Elofsson, 2015; Marks et al., 2011) is a webserver that outputs a score, called evolutionary constraint, which quantifies evolutionary coupling across residue pairs, giving insight into the evolutionary connection between surface and core residues (Hopf et al., 2014). Consurf (Celniker et al., 2013) is another webserver, which provides a conservation score for all residues by using Rate4Site (Pupko, Bell, Mayrose, Glaser, & Ben-Tal, 2002). It gives an evolutionary score by assuming that residues playing important parts in function are conserved across the members of the family of the protein under study. Combining Consurf and EVFold results into an average, for example, yields reliable information on residue conservation.

2.1.3.3 NMA: R and bio3d

NMA has become a widely used method for understanding protein structure–function relations. Normal modes of vibration result from approximating the potential energy surface around the minimum to a harmonic potential and generate an approximate description of the protein motions near the equilibrium state (Brooks & Karplus, 1985; Levitt, Sander, & Stern, 1985; Ma, 2005; Petrone & Pande, 2006). This enables observation of variations specific to the protein–protein interface. In Section 3, an example is provided.

2.2 MD SIMULATIONS IN THE STUDY OF GPCR STRUCTURE, FUNCTION, AND EFFECTORS

MD is a deterministic computational method, which gives information regarding the microscopic properties of biological systems by simulating the time evolution of their atomic coordinates by solving Newton's equations of motion: $a = F/m$. F is

the force exerted on each atom and it is obtained by taking the gradient of the potential, $U(r_1, r_2, \ldots, r_N)$ describing the interaction among all atoms, m is the mass, and a is the acceleration of the atom. In this way, acceleration of the particles in the system can be calculated and used to iteratively update the position and velocity of atoms.

This microscopic information obtained by repeating this cycle for millions of steps is then converted to empirical thermodynamic properties of macroscopic systems such as temperature, pressure, volume, and heat capacity by means of statistical mechanics, which provides rigorous mathematical expressions to relate the distribution and motions of atoms of the N-body system to its bulk properties (Fig. 2).

2.2.1 Force field

A force field is a mathematical expression that describes the dependence of the potential energy of a system, $U(r_1, r_2, \ldots, r_N)$, on the coordinates of its particles in terms of analytical interatomic functions as well as parameter sets and is defined as follows:

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] + \sum_{improper} V_{imp} + \sum_{LJ} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \sum_{elec} \frac{q_i q_j}{r_{ij}}, \quad (1)$$

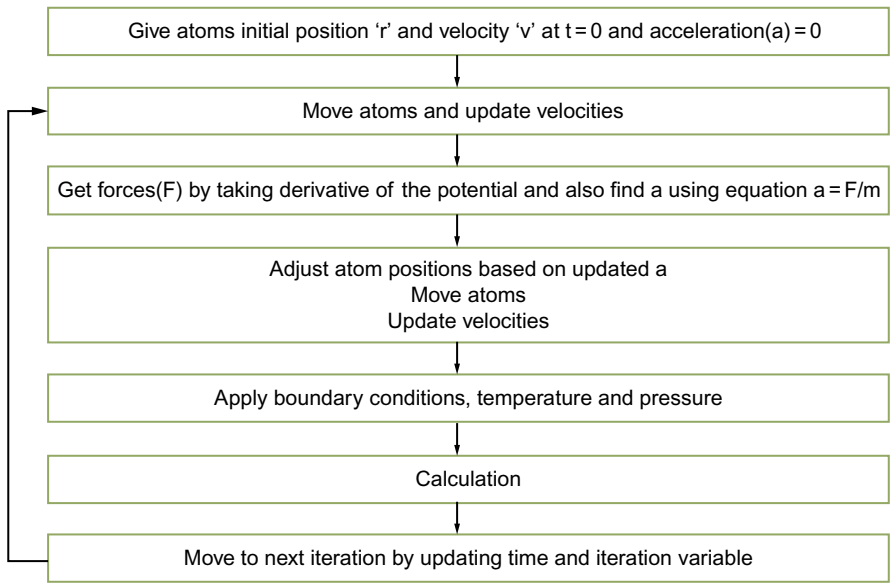

FIG. 2

Scheme of a classic molecular dynamics simulation algorithm.

where the first four terms refer to intramolecular bonded contributions such as bond stretching, angle bending, dihedral torsion, and improper torsion, whereas the last two terms describe the nonbonded contributions like Van der Waals and electrostatic interactions, by means of 12-6 Lennard–Jones and the Coulomb potential, respectively. The parameters of the energy functions can be derived from either experiments or ab initio calculations in quantum mechanics or a combination of both. The commonly used force fields to study protein dynamics are CHARMM, AMBER, and GROMOS (Cornell et al., 1995; Mackerell, Feig, & Brooks, 2004; Monticelli & Tieleman, 2013), which differ from each other by torsional potentials and the way they treat atoms in the system. For further details on the methodology, we refer the readers to related articles (Guixa-Gonzalez, Ramırez-Anguita, Kaczor, & Selent, 2012; Tai, Fowler, Mokrab, Stansfeld, & Sansom, 2008). The accuracy of a force field has to do with its capacity to reproduce correct Boltzmann-distributed conformational ensemble, meaning that conformations of lower energy, which are also captured in X-ray crystallography, are more populated. These stable structures usually correspond to those that govern the function of the protein. The force field should also capture all possible conformations that can be accessed by thermal fluctuations as well, as conformational states involved in receptor activation can be also visited, to a small extent, in the absence of the agonist (Hansen, Vallurupalli, & Kay, 2008). This in turn makes it possible to link statistical fluctuations that are inherent in MD simulations to macroscopic properties of the system by averaging over a sufficient number of independent conformations. The accuracy of these estimates can then be quantified by comparison of the results with experiments providing that simulations are sufficiently converged (van Gunsteren et al., 2006).

In MD simulations of soluble proteins, such as isolated Arrestin or G-protein, a cubic or tetragonal box contains, besides the protein, positive, and negative ions to mimic a physiological salt concentration and water. When simulating a GPCR system or complex, the receptor is embedded in a lipid bilayer, which is surrounded by water molecules. In particular, the force field used to describe properties of lipid molecules in which the receptor is embedded is critical: it has been shown that the membrane lipids, in particular cholesterol molecules, modulate the functional dynamics of the receptor (Khelashvili, Grossfield, Feller, Pitman, & Weinstein, 2009; Patra et al., 2015). Currently, CHARMM force field, in particular CHARMM36, is preferably used to model membrane lipids since there is a consistency between experiments and computation on the lipid properties such as average area per lipid (Nagle & Tristram-Nagle, 2000) and NMR order parameters (Seelig & Seelig, 1974).

2.2.2 Periodic boundary conditions

The aim of MD simulations is to describe a protein in a realistic macroscopic environment, but due to size limitations one is limited to using finite-size systems, where the number of "surface" atoms located in the vicinity of the boundaries constitutes a significant fraction of the overall atoms. In order to minimize the finite-size effects *periodic boundary conditions* (PBCs) are used in MD simulations. The PBCs consist of replicating the simulation box by rigidly translating the system in all directions,

such that replicas (images) completely surround the primary box. The MD simulation is identically replicated in each image. Particles close to the boundary will interact with their images and when leaving the primary box will reappear at the opposite side, which keeps the number of particles constant. In order to keep the number of interacting pairs under control when calculating the energy, and to avoid that periodic images interact with each other, cutoffs are introduced in nonbonded van der Waals interactions, whereas for electrostatic interactions Ewald summation methods are preferred (Darden, York, & Pedersen, 1993). A soluble protein is able to freely diffuse out of the box: due to the PBC it will reenter from the opposite side. In contrast, with GPCRs the receptor is embedded within the membrane, and thus its movement is limited. Yet the PBCs ensure the bilayer to be conserved.

2.2.3 NVT and NPT ensembles

The ensemble of conformations visited by the system during a MD simulation is expected to correspond to a statistical equilibrium distribution at given temperature and pressure conditions. To this end, specific transformations are cyclically applied to rescale particle velocities so that they obey a Maxwell distribution (Berendsen, Postma, van Gunsteren, Di Nola, & Haak, 1984). The dimension of the simulation box (length, width, and height) and the number of particles define the density of the system. If the volume is kept constant, the simulation is carried out in an NVT ensemble. If the volume can fluctuate and be rescaled in order to yield a pressure close to a reference value, the simulation is in the NPT ensemble, which can be done in three ways: *isotropic*, *semiisotropic*, and *anisotropic*. In the first case, the scaling factor is the same in all three directions, whereas in *semiisotropic* coupling, *x*/*y* directions are scaled independently from the *z*-direction. In the *anisotropic* coupling scheme, the scaling factor is applied independently for each of the three axes. The pressure configuration is critical when simulating a receptor in a membrane due to the surface tension of the membrane. Namely, the surface tension of a lipid membrane vanishes at equilibrium in the absence of external stress, and under this condition the membrane exhibits long wavelength undulations. However, these long wavelength fluctuations are suppressed in MD simulations of membranes due to finite-size effects. To alleviate this artifact, the *xy* surface area of the membrane is fixed at tension-free state to maintain the appropriate "area per lipid" value, whereas the pressure along the *z* direction is coupled to a barostat, which can be done by applying the *semiisotropic* coupling, to reproduce the tension-free state of macroscopic membranes (Feller & Pastor, 1996).

2.2.4 Posttranslational modifications in GPCRs

GPCRs as well as their signaling partners undergo posttranslational modifications including phosphorylation, glycosylation, and palmitoylation that affect the functional dynamics, signaling properties (Zheng, Loh, & Law, 2013), and cell–surface expression of receptors (Lanctot et al., 2005). While—in general—glycosylation process occurs at the N-terminus of the receptor, phosphorylation, and palmitoylation instead occur at the cytoplasmic C-terminus of GPCRs. In particular,

palmitoylation is a covalent attachment of palmitic acid moiety to cysteine residues located in the C-terminus of GPCRs, which are conserved in about 78% of 74 GPCRs examined (Probst, Snyder, Schuster, Brosius, & Sealfon, 1992). It has been shown to modulate the membrane depth of C-terminus of either Dopamine 2 (Sensoy & Weinstein, 2015) and Dopamine 3 receptors (Arango-Lievano et al., 2016), and thus its accessibility to signaling partners such as PDZ-domain containing proteins (Sensoy & Weinstein, 2015). In addition, palmitoylation turnover rates have been shown to drastically increase upon receptor activation (Jia et al., 2014) suggesting a possible role of this group in modulating the function of the receptors studied. Therefore, such groups should be properly modeled in MD simulations of GPCRs to reproduce related experimental data. CHARMM force field (Mackerell et al., 2004) has a repository of such moieties in the form of patches that can be attached to the target residue(s) of the receptor.

2.3 LIMITATION OF STANDARD MD SIMULATIONS AND APPLICATION OF ENHANCED SAMPLING METHODS

MD simulations are indispensable computational tools to provide insight at atomic resolution for complementing experimental findings. In particular, MD simulations of GPCRs are usually done in the presence of explicit water and membrane to mimic physiological conditions and this, in turn, causes scaling of the system's size up to the order of several nanometers typically involving ~100,000 atoms or more. In addition, the time step used for the integration in atomistic MD simulation is on the order of 1–2 fs. Under these conditions, the currently available computational resources allow one to reach trajectories spanning length in the order of microseconds or milliseconds at most. Therefore, the current resolution and computational power make it possible—into some extent—to address a number of fundamental questions on receptor activation, linked to subglobal conformational changes and allostery (Jose Manuel Perez-Aguilar, LeVine, Khelashvili, & Weinstein, 2014; Dror et al., 2011; Samuel Hertig & Dror, 2016). However, many relevant conformational events in GPCRs usually occur on time scales of millisecond or higher and most experiments provide the ensemble-averaged structural and thermodynamic properties of the system. Ideally, with MD simulations, the relevant conformational space of the system should be sampled, to provide comparable results. However, the energy landscape of proteins is rugged and characterized by many local minima (Bryngelson, Onuchic, Socci, & Wolynes, 1995; Wolynes, Onuchic, & Thirumalai, 1995) and transitions among them are usually hampered by the presence of high-energy barriers. Consequently, the time-averaged estimates obtained from MD simulations are not comparable to ensemble-averaged properties obtained from experiments and also, simulating conformational transitions involving barriers might become unfeasible. To overcome this problem in systems where ergodicity is hindered by the form of system's energy landscape, enhanced sampling methods have been developed. One possible strategy consists of reducing the number of degrees of

freedom (coarse graining) to effectively simplify the energy landscape and reduce the computation time. Another strategy is based on introducing a bias in the simulation to enhance barrier crossing and exploration of a larger conformational space. In this section, we present two successful approaches in both directions: (1) coarse-grained MD and (2) steered MD.

2.3.1 Coarse-grained (CG) MD

Coarse graining refers to reduction of the number of degrees of freedom, and so the complexity, within the system, by redefining a set of atoms as a single degree of freedom, a bead, and each molecule as a set of beads. One of the most popular CG models is MARTINI (Marrink, Risselada, Yefimov, Tieleman, & de Vries, 2007) that has been used to model dynamics processes in the presence of membrane. In the model, the amino acids are represented by different numbers of beads, ranging between 1 and 5, depending on the size, flexibility, and physicochemical properties. Similarly, fine details of lipid and water molecules are also averaged out, such that, respectively, the lipid head group is presented as beads, lipid tails as bonds, and the water molecule is represented by a single bead. The time step can be increased up to 20–40 fs in the MARTINI force field, instead of 2 fs typical of atomistic simulations. In this way, time and length scales captured in experiments can be achieved. Also, the secondary structure elements such as α-helix and β-strands are constrained in the MARTINI force field and finally, in order to maintain the tertiary structure of the receptor, simulations are performed in the presence of an elastic network, such as *Elnedyn* (Periole, Cavalli, Marrink, & Ceruso, 2009). Therefore, the MARTINI force field may not be well suited for monitoring activation-linked conformational transitions regarding receptor activation, but rather can be used, for instance, to investigate membrane–receptor interactions, in particular, it was successfully applied to model the depth of membrane insertion of amphiphilic Helix-8 of GPCRs in the presence/absence of posttranslational modifications (Arango-Lievano et al., 2016; Sensoy & Weinstein, 2015).

2.3.2 Steered molecular dynamics (SMD)

The method takes inspiration from single-molecule pulling experiment (Grubmuller, Heymann, & Tavan, 1996), where the system is forced toward another known state from an initial equilibrium condition, thus facilitating transitions between different energy minima. Analysis of SMD trajectories gives atomistic insight on the process in question, e.g., types of dominant interactions that mediate the transition. Moreover, the free energy difference between two or more states can also be calculated by means of the Jarzynski's equality (Jarzynski, 1997; Sensoy, Atılgan, & Atılgan, 2017). The method has found a wide range of applications in studying many biophysical processes such as (un)folding mechanisms of proteins (Israelewitz, Baudry, Gullingsrud, Kosztin, & Schulten, 2001; Lu, Israelewitz, Krammer, Vogel, & Schulten, 1998), transportation of compounds through membrane channels (Giorgino & De Fabritiis, 2011; Gwan & Baumgaertner, 2007), and

drug discovery (Colizzi, Perozzo, Scapozza, Recanatini, & Cavalli, 2010). Two groups are defined in a classic steered MD experiment: one is the moving group that is subjected to the force and being pulled throughout the experiment, whereas the other one is the reference group relative to which the moving group is pulled. The pulling experiment can be done in two different ways: by applying *constant force* on the moving group or pulling the moving group with *constant velocity*. Jarzynski's equality connects equilibrium free energy differences, ΔG, between any given two points, which are defined by a parameterized quantity, λ, to the work done through nonequilibrium processes, W. Jarzynski's equality states that the following holds regardless of the speed of the process (Jarzynski, 1997).

$$e^{-\beta \Delta G} = \langle e^{-\beta W} \rangle \tag{2}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann constant, and T is the temperature. By means of repeated SMD simulations with constant velocity, one can accumulate several force plots, one for each pulling experiment, expressing the strength of the interaction between the moving and the reference group as a function of a reaction coordinate $\xi(r)$ that depends on the $3N$-dimensional position r of the system. The force value can be used to calculate the external work as in the following equation:

$$W_{0 \to t} = -kv \int_0^t dt' [\xi(r_{t'}) - \lambda(0) + vt] \tag{3}$$

where $\lambda(t)$ changes with a constant velocity with $\lambda(t) = \lambda(0) + vt$ with $\lambda(0) = 0$.

The bias and errors are estimated using the limit for small number of pulling experiments, N, for perturbations near equilibrium using the scheme developed in Bustamante (2003). Using the average dissipated work as the difference between the average of the work values and the free energy difference calculated through Eq. (2), $\overline{W_{dis}} = \langle W \rangle - \Delta G$, the bias estimate is

$$B = \frac{\overline{W_{dis}}}{N^{\alpha}} \tag{4}$$

where

$$\alpha = \frac{ln\left(2C\overline{W_{dis}}/k_B T\right)}{ln\left[C\left(e^{\frac{2\overline{W_{dis}}}{k_B T}} - 1\right)\right]} \tag{5}$$

Thus the reported free energy is obtained by:

$$\widehat{\Delta G} = \Delta G - B \tag{6}$$

3 METHODS

3.1 CONSTRUCTION AND ANALYSIS OF INTERFACES IN GPCR/EFFECTOR COMPLEXES

In this section, the protocol for building and analyzing D_1R–Gs protein complex is explained step by step. The GitHub repository: https://github.com/IrinaMoreira/gpcr-comparative-analysis contains all the scripts and files utilized in this section. Note that all files included in the scripts should be in the same folder with the scripts or their paths should be adapted accordingly.

3.1.1 Homology modeling (GitHub folder: 1_HOMOLOGY MODELLING)

The first step in the creation of a homology model for D_1R is the identification of appropriate homologs. MODELLER provides an easy and reliable way to do so. The script *1_get_homologs.py* uses the query sequence in *d1r.ali* to search for homologs in the *pdb_95.pir,* a database of nonredundant sequences with known structure. This file should be updated to the latest version and can be found at http://salilab.org/modeller/supplemental.html. This script is prepared to handle the given example but can be easily modified to any other case.

As mentioned in Section 2, special attention should be given to ICL3. If this protein region presents a sequence longer than 20 residues and it has no available template, then a reduced sequence (5–7 amino acids) can be modeled, with an alanine stretch of reduced length (Martinez-Archundia, Cordomi, Garriga, & Perez, 2012) connecting the two extremes. These modifications are introduced at the sequence level in the file *D1R.ali*. The output (*build_profile.ali*) shows that the sequence with the highest identity to the target sequence is the A chain with PDB id 1U19. Upon retrieval of the coordinate file from the PDB, unnecessary chains as well as nonprotein atoms should be deleted. This makes the modeling process simpler and facilitates the visual analysis of both the template and the model. Besides 1U19 (Okada et al., 2004), an alternative sequence was also found for the intramembrane segment of D_1R with higher sequence homology: chain A from the 1HLL PDB file (Chung et al., 2002). For the purpose of this chapter, this should be disregarded since we focus on investigating interactions between GCPR and G protein. Therefore, we have chosen the 3SN6 file (3sn6_A.pdb) as the template, which corresponds to the structure of an active GPCR, namely, β2-adrenergic receptor bound to Gs protein (Rasmussen et al., 2011).

3.1.1.1 Sequence alignment

Sequence alignment can be done by using *2_align_sequences.py* script. From this, two output files are generated: *d1r-3sn6.ali* and *d1r-3sn6.pap*. The first one gives the best representation of our alignment. The file *d1r_colored_fasta.docx* also shows the sequence, color-coded according to the position of TMs, making other regions such as ICLs and ECLs easily identifiable.

3.1.1.2 Homology modeling

By using the *3_build_models.py script*, 100 models can be generated, creating a robust "pool" from which it is possible to select the best model. Evaluating models and selecting the best one can be done using MODELLER object function (MOF) or the DOPE assessment score. Ideally, DOPE should be as low as possible while the MOF score should be as high as possible. A balance between these two rankings can be optimized, leading to the selection of 10 models. Next, a careful analysis using a visualization software such as PyMol (Schrödinger, 2015) should be performed to check if any of the models have their ICL3 loop located in the intramembrane region. As we know the binding crevice should not be occluded to allow the binding of a soluble partner. If any models among the top 10% present this occlusion, they should be discarded, but if all models have plausible ICL3 loop orientations, we should select the one which has its ICL3 most distant from the membrane, to keep steric interference as low as possible upon binding. In our case, upon consideration of all the mentioned criteria, the best model appeared to be *d1r.B99990048.pdb*.

3.1.1.3 Sequence alignment of G-protein using Clustal Omega

When studying the specificity of G protein/GPCR coupling, understanding the similarity between all G proteins is crucial. Hence, a multiple alignment of all G proteins that bind D_1R (Gi1, Gi2, Go, Gslo, and Gssh) can be performed, using for instance the Clustal Omega webserver (Sievers & Higgins, 2014), available at: http://www.ebi.ac.uk/Tools/msa/clustalo/. As an input, protein sequences should be given in FASTA format (see examples: *D1R.fasta*, *Gi1.fasta*, *Gi2.fasta*, *Go.fasta*, *Gslo.fasta*, and *Gssh.fasta*) and submitted together, separated by full line breaks. Optionally, a file can be provided in the same format. The results can be downloaded by clicking "Download Alignment File." Clustal format files align the sequences presenting gaps (−) to match as many residues as possible. To facilitate the introduction of this alignment into the AT, a script is available (*clustal_toalign.py*), which finds all Clustal format files and converts them to an AT. An example format of AT is presented in the 1_HOMOLOGY MODELLING/SEQUENCE ALIGNMENT folder (*Aligned Table.xls*).

3.1.2 Structure refinement and complex docking. GitHub folder: 2_STRUCTURE DOCKING AND REFINEMENT

3.1.2.1 Structure alignment and refinement

If an experimentally determined structure for a GPCR–G protein complex is available, the simplest way to get a starting complex is to align the monomers to be studied with the ones in the complex using, e.g., the visualization software PyMol (Schrödinger, 2015). After loading both experimental and modeled structures to the PyMol session, structure alignment can be performed ("Action" → "align" → "to molecule" → experimentally determined structure).

Following this, HADDOCK's refinement interface can be applied (available at: http://haddock.science.uu.nl/services/HADDOCK2.2/haddockwebserver-refinement.html upon registration). To use this interface, "Expert" access should be requested by

email. Instructions for the structure submission are explained in detail in the webpage. The best structure can then be retrieved from the output page, which lists the refined interface models with the highest score. Example of output structures are presented in the 2_STRUCTURE DOCKING AND REFINEMENT/OUTPUT folder in the GitHub repository, while example input structures are present in the STRUCTURE DOCKING AND REFINEMENT/INPUT folder. If no template for the complex is available, the Easy web interface (available in: http://milou.science.uu.nl/services/HADDOCK2.2/haddockserver-easy.html) implemented in the HADDOCK webserver provides a reliable starting structure by using knowledge on interfacial residues to improve the search and scoring algorithms of the docking approach.

3.1.3 Structural features. GitHub folder: 3_STRUCTURAL_FEATURES

For the remainder of this section, it is necessary that all PDB files contain the two chains, one for the GPCR and other for its effector.

3.1.3.1 VMD: SB

Loading a structure into VMD is done using the "New Molecule..." option under the "File" menu. Following this, using the SB extension for VMD ("Extensions" → "Salt-Bridges") it is possible to set a desired cutoff distance for nitrogen–oxygen interactions. The output is given in a file for each SB, listing interacting residues and their distance in Å units. Only interchain residues should be considered, when analyzing the protein–protein interface.

3.1.3.2 CoCoMaps: HB, ASA

Input files of CoCoMaps (available at: https://www.molnac.unisa.it/BioTools/cocomaps/) should be prepared according to the instructions, located under the OPTION 2, and starting with a PDB file. The specified chains under "Chain Molecule 1/2" should contain the chains forming the complex (GPCR and its effector). Even though the output file contains a high amount of information, the tables to be stored are: (i) "Interaction overview," (ii) "H-Bonds Table" and the three ASA tables: (iii) "ASA Table," (iv) "ASA Table for Molecule 1," and (v) "ASA Table for Molecule 2." Values for (i) and (iii) can be stored in a table such as CoCoMaps_Interface.csv, available in the GitHub repository, while values for (ii), (iv), and (v) can be stored in the AT.

3.1.4 Evolutionary features. GitHub folder: 4_EVOLUTIONARY FEATURES

3.1.4.1 EVFold

EVFold (available at: http://evfold.org/evfold-web/newprediction.do?) requires only the sequence of the protein to be submitted and no structural information is requested. Sequences of GPCR and its effector should be submitted separately. The output includes for each monomer MSAs and 3D structure prediction, but we focus here on the EC hotspot table. This can be accessed in the "Evolutionary Constraint (EC) Table" menu by clicking the "View EC Hotspot Table." This table has several scores (cumulative strength, EC strength, and conservation), of which the

second—EC strength—is the one to be stored in an AT. This webserver calculates MSAs for each input sequence. This step is computationally intensive, as EVFold aligns hundreds of thousands of sequences.

3.1.4.2 Consurf

Consurf, like EVFold (http://consurf.tau.ac.il/2016/), requires the protein sequence only. On the website there is a straightforward pipeline to follow. The default parameters should not be changed by nonexpert users to execute the MSA job. In the results, the "Amino Acid Conservation Scores, Confidence Intervals and Conservation Colors" section contains a table with the normalized conservation score, "SCORE (normalized)." For GPCR–effector interface analysis, the information retrieved from Consurf conservation scores can be combined to the one provided by EVFold EC strength, providing a consensus view on EC.

3.1.5 Comparative NMA. GitHub folder: 5_COMPARATIVE NORMAL MODE ANALYSIS

By combining R, a programming language with increasing popularity among computational biologists (Team, 2008) and bio3d (Grant, Rodrigues, ElSawy, McCammon, & Caves, 2006), which is an R package that provides a comparative NMA platform, a highly customizable way for comparing normal modes is available. A custom R script with instructions can be found in the GitHub repository (NMA.R). The output lists the interfacial residues in cvs files. In addition, graphs representing fluctuations for all residues with particular focus on the interface residues are also provided. To utilize this script, it should be placed in a folder, which contains only PDB files with two chains, the first one for GPCR and the second one for its corresponding effector. The PDB files should be named as "GPCR_partner.pdb"—e.g., "D1R_Gi1.pdb"—in order for the script to create a graph with accurate information (these are case sensitive). The script is executed calling "Rscript NMA.R" in the same folder.

The interfacial residues provided in the csv output files are ordered alphanumerically. And can be identified by consulting the file "all_pdb.csv," which contains general information on each PDB (monomer names and chains). The final fluctuation graph is constructed based on the dynamical information stored in "fluctuation.csv" and is named after the first monomer on the input file name. For example, if the input file names are "D1R_Gi1.pdb," "D1R_Gi2.pdb," "D2R_Gi1.pdb," and "D2R_Gi2.pdb," the output graphs containing the fluctuations of each effector when bound to D_1R or D_2R are named "D1R.png" and "D2R.png," respectively, and the fluctuations of each dopamine receptor when bound to different partners are named "Gi1.png" and "Gi2.png." Residue numbering and interfacial residues for all graphs are relative to the alphabetically first GPCR or partner. In the example presented, the numbering is relative to Gi1.

By combining this information with the sequence alignment (see Section 3.1.1), we can identify important residues in the interface by analyzing the variations for these amino acids. Since we are expecting small variations to occur in the interface,

it is recommended to adjust the scale for the *y* scale, line 88 of the script. For G proteins, for example, it is interesting to see that, when analyzing interface residues and neighboring residues, activating and inhibitory G proteins have clearly different residue fluctuations in the first interfacial residue window (28–32) and in the ones closer to the end (192–321 and 336–354), as can be seen on the D1R.png figure in the GitHub repository.

3.1.6 Comparative structural and evolutionary analysis. GitHub folder: 6_COMPARATIVE ANALYSIS

While the last step of the previous section features a built-in method for comparative analysis, comparing the rest of the data is necessary to reveal important patterns or investigate specific interactions. This requires a higher amount of "hands-on" approach than the previous step as we need to search across the whole collected information by means of the AT.

In order to transform our data into more accessible information, a filter is defined through the general conservation score (GCS), obtained by averaging all residues' Consurf and EVFold scores, only residues featuring a GCS scores should be considered. This highlights residues that are important for the overall structure while being simultaneously conserved. If the highlighted residues have any interfacial values associated to them (SB, HB, or ASA), we can corroborate the hypothesis that these structural features are of high importance to the protein–protein interface.

In order to comprehensively characterize all GPCR–partner interfaces, the remaining information from CoCoMaps, not used to build the AT, should be considered. That includes, e.g., the interface energy/area, number of polar/apolar residues, and hydrophobic/hydrophilic interactions. Moreover, GCS defined above can be combined to this structural information to provide a reliable picture of our data. Two examples (one for the CoCoMaps interfacial energy and one for the CoCoMaps ASA values) can be seen in the Comparative Analysis Example.xlsx in the GitHub repository.

3.2 SETUP OF A MD SIMULATION OF A GPCR IN THE MEMBRANE (IN ATOMISTIC REPRESENTATION)

Once the structure of the complex has been generated and analyzed, the next level of conformational analysis can be to evaluate the equilibrium and the time-dependent properties of the system by means of MD simulations. Here, the main steps for setting up of a MD simulation of a GPCR are summarized. The following tools are required for this tutorial:

- VMD, available at http://www.ks.uiuc.edu/Research/vmd
- NAMD, available at http://www.ks.uiuc.edu/Research/namd/

3.2.1 Retrieving and examining the structure of the GPCR of interest

If the 3D structure of the GPCR or complex under study is available and deposited in the PDB, VMD can be used to retrieve a PDB file. To do so, the four-letter code of the protein is written in the "File Name" text entry of the "Molecule File Browser window" and upon pressing the "Load" button VMD will automatically download the file. As an alternative, the structure should be provided by a homology modeling protocol as discussed above.

As GPCRs are in general difficult to crystallize due to presence of highly flexible regions in the receptor such as extracellular and intracellular loops, experimental data can contain stabilizing mutations, or specific antibody fragments at these regions, or they can be entirely removed. Importantly, the details of such modifications are stored in the PDB file. Therefore, this must be examined carefully and the native receptor must be restored before starting MD simulations, either introducing mutations or via homology modeling, as discussed in the previous section.

The next step consists of determining which constituents of the system originally present in the PDB file will be included in MD simulations. In general, lipids molecules and ions are removed, whereas crystal water molecules are kept as they are important for the functional dynamics of the receptor.

Subsequently, the protonation states of ionizable amino acids such as histidine, arginine, lysine, glutamic, and aspartic acid need to be modeled, which can be done using a software such as propKa (Olsson, Sondergaard, Rostkowski, & Jensen, 2011). Lastly, disulfide bonds, if applicable—and posttranslational modifications must also be added to appropriate residues since these groups modulate the stability and the dynamics of, respectively, the extracellular and the cytoplasmic parts of the receptor. In the presence of a homology model, such as the complex built in the previous section, the same protocol can be applied. As an alternative to the package psfgen of VMD, the web interface available at: www.charmm-gui.org/?doc=input/pdbreader provides a tool to process an input pdb file and introduce a number of posttranslational modifications to generate the receptor structure .psf file required for the subsequent steps.

3.2.2 Placing the GPCR into a membrane

Membrane proteins should be simulated in their native-like environment, in particular, in the presence of lipids as found in vivo conditions. The VMD membrane builder plugin automates the preparation of a complete membrane by replicating preequilibrated patch of membrane and water, and then trimming it if needed. VMD provides however only POPC and POPE membranes, but several other types of lipids are supported by the CHARMM force field and can be obtained from http://www.charmm-gui.org/?doc=input/membrane. Once the membrane is formed then the receptor is aligned properly with respect to the membrane. To this end, first the membrane will be aligned with its center of mass by using the following command line in VMD:

```
set popc [atomselect top all]
$popc moveby [vecinvert [measure center $popc weight mass]]
```

For the GPCR, a specific region of the receptor should be chosen of which center of mass will be used to rotate the receptor about the z-axis in order to align roughly the top end with the x- and y-axes by using the following command lines in VMD:

```
set GPCR [mol new GPCR.psf]
set GPCR_all [atomselect $GPCR all]
set spec_GPCR [atomselect $GPCR "specific region of the receptor"]
$GPCR_all moveby [vecinvert [measure center $spec_GPCR weight mass]]
$GPCR_all move [transaxis z -25]
```

The next step is to make room for the receptor in the membrane so that it does not overlap with any lipid molecules. This can be done by marking the overlapping atoms by means of the beta column of the PDB file. First, the beta column of all the atoms is set to zero in VMD by the following command lines:

```
set all [atomselect top all]
$all set beta 0
```

Second, appropriate selections are made to mark the lipids whose phosphorus atoms overlap with the receptor and then mark the rest of the lipids within a certain cutoff distance of the protein.

```
set seltext1 "$POPC and same residue as (name P1 and z>0 and abs(x)<15
and abs(y)<15)"
set seltext2 "$POPC and same residue as(name P1 and z<0 and abs(x)<10
and abs(y)<10)"
set seltext3 "$POPC and same residue as (within 0.6 of protein)"
set sel1 [atomselect top $seltext1]
set sel2 [atomselect top $seltext2]
set sel3 [atomselect top $seltext3]
$sel1 set beta 1
$sel2 set beta 1
$sel3 set beta 1
set badlipid [atomselect top "name P1 and beta > 0"]
set seglistlipid [$badlipid get segid]
set reslistlipid [$badlipid get resid]
```

Here, "POPC" and "P1" refer to the POPC lipid and the phosphorus atom, respectively. In addition, the values given above are specific for a system and so they should be adjusted properly depending on the size of the system in question.

The membrane patch used to embed the protein also includes water molecules which solvate the head groups of lipids. Among them, some water molecules may overlap with the receptor, and therefore have to be removed from the system as well. This can be done in the following in VMD:

```
set seltext4 "(water and not segname WCA WCB WCC WCD WF SOLV) and same
residue as within 3 of((same residue as (name P1 and beta>0)) or
protein)"
set seltext5 "segname SOLV and same residue as within 3 of lipids"
```

```
set sel4 [atomselect top $seltext4]
set sel5 [atomselect top $seltext5]
$sel4 set beta 1
$sel5 set beta 1
set badwater [atomselect top "name OH2 and beta > 0"]
set seglistwater [$badwater get segid]
set reslistwater [$badwater get resid]
foreach segid $seglistlipid resid $reslistlipid{
delatom $segid $resid
}
foreach segid $seglistwater resid $reslistwater{
delatom $segid $resid
}
```

3.2.3 Solvation and ionization of the system

The VMD solvate plugin places the protein in a box of water of a specified dimension and removes water molecules that are put inside of the lipid membrane. First, in order to measure the thickness of the water layer **minmax** option of the **measure** command can be used as in the following in VMD:

```
set water [atomselect top water]
measure minmax $water
```

The size of the water box should be of similar size in the *xy*-plane. However, for nonequilibrated membranes the *xy*-plane of the water box should be slightly smaller since the lipid molecules in such systems tend to shrink. Once the dimension of the water layer is determined the system can be solvated by using the following command lines in VMD:

```
package require solvate
solvate X.psf X.pdb -o X_solvated -b 1.5 -minmax{{-38 -38 -39}
{39 39 50}}
```

Here, X refers to the name of the system file. –b option is used to remove atoms within 1.5 Å of the receptor. The water molecules put inside the lipid bilayer and around the protein can be removed by using the following command lines in VMD:

```
set all [atomselect top all]
$all set beta 0
set seltext "segid WT1 to WT99 and same residue as abs(z) < 25"
set sel [atomselect top $seltext]
$sel set beta 1
set badwater [atomselect top "name OH2 and beta > 0"]
set seglist [$badwater get segid]
set reslist [$badwater get resid]
foreach segid $seglist resid $reslist{
delatom $segid $resid
}
```

Living organisms are under tight regulation to maintain the concentration of ions inside and outside of the cells. Therefore, in order to mimic physiological or corresponding experimental conditions, the ionic concentration of the simulation box should be adjusted properly. This can be done by using the Autoionize plugin of VMD, which creates a specified ion concentration of either KCl or NaCl by transforming randomly selected water molecules into ions.

3.2.4 Running a simulation of a GPCR embedded in the lipid membrane

When the membrane patch is not equilibrated—like the one provided by the plugin of VMD used above—it is recommended to perform a simulation first where the components of the system (ions, water, receptor, and lipid head groups)—except lipid tails—are fixed in order to induce relevant disorder of a fluid bilayer. Once this is done, the next step is the energy minimization, to guide the system to the nearest local energy minimum. This should be followed by an equilibration step with the protein atoms harmonically constrained in order to allow the system constituents to relax around the protein structure. This can be done by including the following lines in the NAMD simulation run file.

```
constraints on
consexp 2
consref X.pdb
conskfile X.cnst
conskcol B
tclforces on
set waterCheckFreq 100
set lipidCheckFreq 100
tclForcesScript keep_water_out.tcl
```

The first parameter set is used to impose harmonic constraints on the receptor. In particular, *consexp* describes the order of the function that is used to impose the constraints. The identification of the atoms to which harmonic constraints are applied is done by using the Beta column (*conscol B*) of the corresponding PDB file. These constraints let the system constituents including lipids, water, and ions to equilibrate around the receptor.

The second parameter set, which is described in a Tcl script named *keep_water_out.tcl* (see https://sassieweb.chem.utk.edu/training/aps_2016/files/lab_VIII_membrane_builder.pdf) is used to prevent hydration of the membrane–receptor interface during the equilibration step.

```
set all [atomselect top "all"]
$all set beta 0
set prot [atomselect top "protein"]
$prot set beta 1
$all writepdb kcsa
popcwi.cnst
exit
```

The commands above are used to create a file named *X.cnst* that contains zeros in the Beta field of all atoms except those that belong to the receptor, which contain 1. The latter value corresponds to the spring constant "*k*" of the applied harmonic constraint in kcal/mol/Å^2. Subsequently, this step will be followed by another equilibration phase with the receptor released, and, then the system becomes ready for the production run to accumulate data of interest.

3.3 SETUP OF A MD SIMULATION OF A GPCR IN THE MEMBRANE (IN CG REPRESENTATION)

The tools and the scripts used in this tutorial are given below:

- Gromacs (http://www.gromacs.org/)
- Martinize.py can be downloaded at http://www.cgmartini.nl/index.php/tutorials-general-introduction/proteins#membrane-protein
- Insane.py can be downloaded at http://www.cgmartini.nl/index.php/tutorials-general-introduction/proteins#membrane-protein
- Python (https://www.python.org/downloads/)

An interactive flowchart is provided in the website of Martini (http://www.cgmartini.nl/index.php/tutorials-general-introduction/flowchartfile) to guide the user for an effective CG simulation of the system of interest. To do so, it is required to transform the system constituents (receptor, water, membrane, and ions) into the CG representation. As to the membrane this can be done in three ways: (1) the Martini has a wide repository of lipid molecules. If the lipid studied is available it can be downloaded from http://www.cgmartini.nl/index.php/force-field-parameters/lipids along with the topology file, (2) if there are similar lipid molecules available in the repository, the type of the lipid can be appropriately changed since they are modular molecules, and (3) the bilayer can be self-assembled from scratch, but this time the following properties should be checked as well: area per lipid, bilayer thickness, P_2 order parameters of bonds, and lateral diffusion. In order to prepare more complex and larger bilayers it is more convenient to start with a bilayer that is close to equilibrium, which can be done by concatenating/altering preformed bilayers or by using a bilayer formation program such as *insane.py* (INSert MembrANE). It generates bilayers by distributing lipids over a grid. The program uses two grids, one for the inner and the other for the outer leaflet, and distributes the lipids randomly over these grid cells according to the specified ratios. In addition, solvent and ions can also be added using a similar grid protocol which distributes them over a 3D grid.

```
python insane.py -l DPPC:4 -l DIPC:3 -l CHOL:3 -salt 0.15 -x 15 -y 10 -z 9 -
d 0 -pbc cubic -sol W -o X.gro
```

Here, DPPC, DIPC, and CHOL refer to specific lipid molecules. –salt option determines the molarity of the salt used to neutralize the system, whereas –x, y, and z determine the number of lipid molecules to be placed along the axes. This command

line will generate an initial configuration file X.gro, which should subsequently be minimized and equilibrated.

As to the coarse graining of the receptor, the following command line can be used:

```
martinize.py -f X-atom.pdb -o 0.top -x 1X-CG.pdb -dssp -p backbone -ff
martini22
```

When using the *-dssp* option one needs the dssp binary (Kabsch & Sander, 1983), which can be downloaded from http://swift.cmbi.ru.nl/gv/dssp/. The program determines the secondary structure classification of the backbone of the receptor from the structure. As an alternative, one may prepare a file with the required secondary structure and feed it to the script as shown below:

```
martinize.py -f X-atom.pdb -o 0.top -x X-CG.pdb -ss <YOUR FILE> -p
backbone -ff martini22
```

Once all the system constituents are obtained in the CG representation, then the next step would be to insert the receptor into the membrane, which can be done as follows:

```
insane.py -f X.gro -o system.gro -p system.top -pbc square -box 10,10,10
-l DPPC
-center -sol W
```

This command should build up a complete system of DPPC bilayer of 10 nm, in which the receptor is centered. In addition, the whole system is solvated and ionized in the CG representation.

3.3.1 Reverse transformation: Converting the CG representation of a system into the atomistic one

Although CG simulations give access to larger time and length scales, the loss of atomic resolution can limit the questions that can be addressed. Therefore, methods that provide reintroducing atomic details in the CG representation can help overcome this problem. In the following, the steps required to convert the CG representation of a system to the atomistic one are summarized. For this purpose, a modified version of GROMACS, which can be downloaded from http://www.cgmartini.nl/index.php/tools2/36-downloads/tools/113-rt, will be used that allows one to generate a fine-grained (FG) structure from CG beads (Rzepiela et al., 2010) by means of a simulated annealing algorithm. To achieve this, additional information is required in the topology file at the FG level in a section called [mapping]. The topology and input files needed for this transformation can be downloaded from http://www.cgmartini.nl/index.php/tools2/36-downloads/tools/113-rt/rev_trans.tar.gz. First, one needs to compile and source the modified version of Gromacs as follows:

```
source /where-ever-you-installed-it/gromacs-3.3.1/bin/GMXRC
export GMXLIB=/where-ever-you-installed-it/gromacs-3.3.1/share/
gromacs/top
```

Then, the FG *fg.top* file should be modified such that the number of water and lipid molecules is correctly obtained from the CG representation. (One CG water corresponds to four FG water molecules.) By using the following command line an input atomistic structure file for a simulated annealing run is prepared.

```
g_cg2fg -pfg fg.top -pcg lipid_cg.top -n 1 -c cg.gro -o fg.gro
```

By using grompp command in GROMACS one can create a topol.tpr file to be used with the mdrun in the next step. Then, perform a simulated annealing run by the following command line:

```
mdrun -coarse cg.gro -v
```

By changing the number of simulation steps and simulated annealing time parameters in the .mdp file, which is stored in rev_trans.tar.gz, one can adjust the level of the resolution of the FG structure of the system.

3.4 ANALYSIS OF MD TRAJECTORIES

The typical questions that are addressed when running a standard, unbiased MD simulation have to do with the prediction of conformational states and dynamic modulation to be compared to experiments, as well as model functional mechanisms at the molecular level. The methods can include both structural and dynamical analyses.

Structural analyses address the question of characterizing the equilibrium average structure and in general the conformational preferences of the protein structures or of subdomains under given conditions or in response to perturbations, such as mutations, biased ligands, or in the context of complexes, such as GPCR–arrestin or GPCR–G binary complexes. In particular the structures can be monitored by inter atomic distances, hydrogen bonding patterns, rotation angles, and other geometrical measures (see for instance, Arango-Lievano et al., 2016; Jose Manuel Perez-Aguilar et al., 2014). In contrast, dynamic analyses address the modulation of protein flexibility, of fluctuations and correlated motions and are suited for analyzing protein allosteric properties.

Currently available simulation and visualization packages offer a number of tools to perform a wide range of analyses that can be integrated into protocols. A powerful suite of analysis and visualization tools is again provided by VMD (Humphrey et al., 1996). Also, the GROMACS simulation toolkit (Berendsen, Spoel, & Drunen, 1995; Hess, Kutzner, van der Spoel, & Lindahl, 2008) offers a wide range of analysis methods. However, depending on the MD package that was used to produce the trajectory, the format for trajectory and topology files might need to be converted into the GROMACS compatible formats, .trr or.xtc or into .pdb. The latter is however not recommended due to the rapidly exploding file size for a nonbinary format. To this end, a useful tool, either as a standalone package or as VMD plugin, is the program catdcd (http://www.ks.uiuc.edu/Research/vmd/plugins/catdcd/), which allows one to convert trajectories among a series of different MD-based formats. Also, in order to obtain a GROMACS compatible topology .top file, which is needed to generate the

structure .tpr file, the pdb2top tool has been developed and can currently be retrieved from the GROMACS development repository or GitHub. Most analysis tools provided by GROMACS however do not require a structure file as an input, in most cases a .pdb file will be sufficient (see Manual at www.gromacs.org).

Any analysis and measure require specifying the single atoms or the group of atoms on which it has to be performed. VMD plugins typically offer selection windows where this information can be inserted. In GROMACS, the information on all possible groups and selections, to which one refers to when invoking an analysis tool, is stored in an index file, which can be generated and edited (including adding new selections of atoms or residues) with the command make_ndx (see Manual for details).

3.4.1 Convergence

The production run of a MD simulation aims at sampling the equilibrium conformation, or ensemble of conformations, of a protein system under the given conditions of temperature, pH, bound ligands, or protein interacting partners, in order to investigate the most probable structures in solution and, in the case of a GPCR, embedded in the membrane. Depending on the starting conformation, the equilibration might include a variably long time interval after the simulation has started, where the atom positions evolve toward a relaxed configuration corresponding to the overall potential energy and kinetic energy equilibrium value. Referring to the energy landscape model (Okazaki, Koga, Takada, Onuchic, & Wolynes, 2006; Wolynes et al., 1995), one can imagine that the protein structure at the beginning of the simulation lies in the vicinity of a global free energy minimum. Therefore the simulation travels toward the basin of the minimum, and once it has reached, there is expected to remain for most of the time.

A typical measure to monitor the convergence of the simulation in this sense is therefore the root-mean-square deviation (RMSD) of the protein structure relative to a reference conformation, like for instance, the starting one. The RMSD is defined as:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i)^2} \tag{7}$$

where d_i is the distance between the position of atom i in the structure under consideration and the position of the same atom in the reference conformation. It describes the average distance between two conformations, based on the mean square distance between the two different positions of each atom. Very often only the Cα atoms are considered. Normally, algorithms calculating the RMSD apply a rigid rotation/translation to the protein to maximize the overlap with the reference structure (Maiorov & Crippen, 1995) and minimize the RMSD value.

During the equilibration phase the RMSD typically increases up to a plateau, corresponding to the stage where all protein atoms progressively relax to a thermally equilibrated conformation. Depending on the size of the system, this phase may include several tens of nanoseconds and has to be discarded for subsequent analysis.

Beyond that point, the evolution of the RMSD ideally depends on the conformational landscape that is sampled. In the presence of a single free energy minimum one can expect the conformation to stay close to an average structure and the RMSD to fluctuate around a constant value. In contrast, in the presence of globally different interconverting structures separated by energy barriers of the order of few ~KT, like for instance a loop assuming two alternative arrangements, one would like to ideally see multiple transitions between the states, which correspond to stepwise variations of RMSD. Sampling limitations usually prevent this from happening, yet the typical behavior of the RMSD evolution might show changes all along the simulation. Ideally, the assessment of the convergence of a simulation is crucial to determine the reliability of the conclusions and one has to keep in mind that conformational transitions might have intrinsically different timescales and hence require different sampling intervals before reaching equilibrium.

Regardless of a stable RMSD profile even for long-term simulations in the order of microseconds, even in the presence of converged parameters like potential energy or density, e.g., of lipids in the membrane, it is recommended to assess convergence using more accurate methods (such as Chodera, 2016). Also, the use of replicas, independent trajectories generated with the same starting conditions but with a different starting velocity distribution, provides an indirect means of validating the robustness of the results, as long as the same conformations occur in the replicas, which indicate a sufficient exploration of the energy landscape.

Once the dataset has been generated after discarding the equilibration phase from the production, one can proceed with further analysis.

The RMSD can be calculated for instance with a plugin of VMD, called RMSD trajectory tool (see http://www.ks.uiuc.edu/Research/vmd/plugins/rmsd/) and in GROMACS using the command g_rms (see GROMACS manual at www.gromacs.org for examples and detailed explanation).

3.4.2 Methods for structural analysis of MD simulations

3.4.2.1 RMSD

This quantity is used as a basic criterion to monitor the structural evolution during MD, not only to assess convergence but also to highlight conformational transitions and their time dependence. Transitions are typically visible as sharp increases in the RMSD relative to the starting structure. However, the amount of information that can be retrieved from the RMSD value is limited, due to its poor resolution.

3.4.2.2 Structural clustering

The clustering analysis is used to classify the snapshots visited along a MD trajectory according to their conformational similarity and divide them into groups of conformations that are structurally alike.

This kind of analysis allows one to highlight the most frequent conformations assumed by the receptor, as well as to assess the existence of multiple significant structural arrangements. It is particularly useful to highlight structural modulation induced by different ligands on the same receptor (examples can be found here).

Finally, it can be a suitable method to map along the dynamics transitions between different conformational ensembles.

The similarity measure between two structures has to be defined as a distance (i.e., a nonnegative number, equal to zero when measuring the distance of a structure to itself, and satisfying the triangular inequality, i.e., the distance between two points cannot exceed the sum of the distances between each one of them and a third conformation). A good definition of distance between two structures is the RMSD introduced in the previous section. Many clustering algorithms exist, which differ in the criteria that are used to group the structures: some of them find the optimal decomposition into a given number of clusters (K means) made of structures close to a center conformation (Hartigan & Wong, 1979), some other focus on the local density of data points (Rodriguez & Laio, 2014; Zen, Carnevale, Lesk, & Micheletti, 2008). GROMACS offers a toolkit of different clustering algorithms that are invoked with the command g_cluster (see GROMACS manual for a detailed explanation of the options). Among others, one widely used algorithm provided by the g_cluster tool is the gromos method introduced in the paper of Daura et al. (1999). It is based on a cutoff that is used to discriminate neighbors (pairs of structures at a distance below the cutoff). The structure having the maximum number of neighbors is identified to define the first cluster, made of the set of all neighbors and their neighbors, which is then taken out of the pool of structures. The process is repeated by defining new clusters until there are no structures left.

The cutoff is the critical parameter that has to be chosen in order to produce a meaningful cluster decomposition. A reasonable choice might depend on the number of amino acids of the system and on the size scale of the structural changes one is interested in discriminating, but typically ranges between 1 and a few Angstroms. A too short cutoff will result in a large number of clusters (in the order of hundreds), each one containing a small percentage of snapshots. This will cause the decomposition to be noisy and not very informative. In contrast, a too long cutoff will result in a poor decomposition with all structures belonging to a single cluster.

This clustering method can be successfully applied either considering the full system, typically calculating the Cα atoms only, or to a subset of residues whose structural arrangement is critical to understanding protein rearrangements.

As an example, two alternative cluster decompositions of a MD trajectory of human Arr3 (Sensoy, Moreira, & Morra, 2016) are shown here (Fig. 3). In the first decomposition, the full set of Cα atoms is used, with a cutoff of 2 Å. This analysis generates 35 different clusters, with the most populated cluster representing only 13% of the dataset. This is due to the presence of the long flexible loop comprising residues 345–395 (protein numbering: 1–403), which explores a wide set of conformations, inducing a strong structural variability (in Fig. 3A, the RMSD matrix is shown in the upper diagonal, with the cluster transitions in the lower diagonal).

In the second attempt of decomposition, the segment corresponding to the long interdomain loop residues is removed from the group of atoms used in the cluster analysis. The same command, with the same cutoff on the subset of Cα atoms excluding the loop, yields a smoother decomposition into nine clusters. Here cluster 1, which is identified as the preactivated ensemble (see Fig. 3B and C), occupies

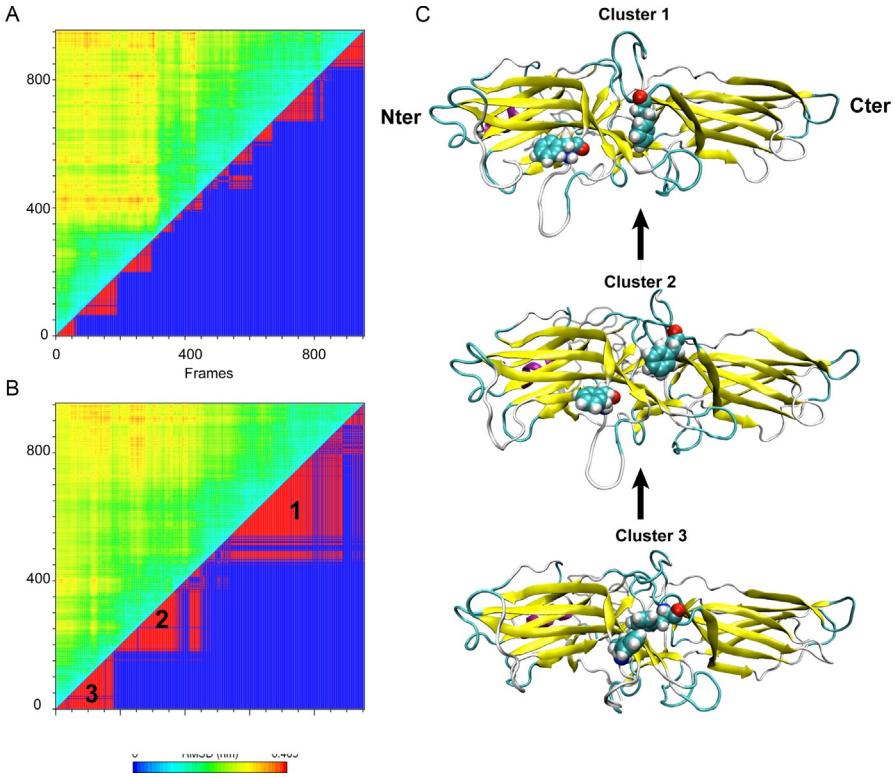

FIG. 3

Clustering of a MD trajectory of Arrestin 3 undergoing activation. (A) Matrix showing the RMSD values of any two snapshots including all Cα atoms (*upper half*) and cluster transitions (*lower half*; *red*=same cluster, *blue*=different cluster, see GROMACS manual for details). (B) Same as (A) but calculating the RMSD and clustering excluding loop residues 351–400. (C) Structural evolution from cluster 3 to 1 showing disruption of the aromatic core (residues 245, residue 76). For details on the MD simulation, see Sensoy et al. (2016).

roughly 39% of the simulation time, whereas clusters 2 and 3, with 25% and 18% occupancy, respectively, show an intermediate step and the starting ensemble. In particular, the unfolding of the small helix discussed in Sensoy et al. (2016) is clearly visible in the transition $3 \rightarrow 2 \rightarrow 1$. This example clarifies the importance of tuning clustering cutoff and defining the most appropriate set to address the description of structural evolution in a MD trajectory.

3.4.3 Methods for dynamical analysis of MD simulations

Dynamic measures provide information on protein flexibility and on correlated motions, both at the local level and at the global level. The local flexibility modulation can give insight into the response to a ligand in the vicinity of the binding site or at a

protein hinge position, the global correlations can give insight into protein functional dynamics, subdomain motions, and allosteric phenomena (Jose Manuel Perez-Aguilar et al., 2014). All these aspects are of critical importance in the study of the molecular mechanisms underlying GPCRs function.

3.4.3.1 Root mean square fluctuations (RMSF) and distance fluctuations

The RMSF relative to the average position can be defined for each atom of the protein, to some reference conformation. With GROMACS, the instruction g_rmsf yields a profile of the fluctuations for the groups of atoms whose selection is prompted by the program, once the trajectory has been aligned to structure.pdb.

The alignment is a critical step, particularly in the presence of significant subdomain motions during the trajectory. The –fit label controls that the trajectory be aligned to the given reference structure, but can be switched off if one prefers to align the trajectory to more specific subgroups of atoms, or single domains in multidomain proteins, for instance by making use of the RMSD trajectory tool plugin of VMD (http://www.ks.uiuc.edu/Research/vmd/plugins/rmsdtt).

The limitation of the RMSF measure emerges when calculating the fluctuation profile for proteins undergoing rigid motions of subdomains: in this case it is typically not possible to discriminate between flexible hinges and rigid parts, as they all give rise to high fluctuations moving together. To highlight flexible hinges, as well as to discriminate between locally flexible regions and rigid subdomains, one can evaluate distance fluctuations instead. Given a system of N atoms, the distance fluctuations map is a $N \times N$ matrix of entries A_{ij}, where each entry is calculated over the trajectory as:

$$A_{ij} = \langle \left(d_{ij} - \langle d_{ij} \rangle \right)^2 \rangle \tag{8}$$

where d_{ij} is the (time dependent) distance between the Cα atoms of amino acids i and j and the brackets indicate time average over the simulation. Each matrix entry reports on fluctuation of the interresidue distance in the corresponding residue pair. Lower distance fluctuation values correspond to a higher internal coordination between the residues (local rigidity). Matrix regions showing relatively low values identify protein subdomains that move together (in coordination) while undergoing structural fluctuations (see Chiappori, Merelli, Colombo, Milanesi, & Morra, 2012; Sensoy et al., 2016).

3.4.3.2 Essential dynamics analysis

It is hypothesized that the most relevant motions connected to protein function are collective breathing motions, occurring on the longest timescales (in comparison to the simulation time), and involving many degrees of freedom (Amadei, Linssen, & Berendsen, 1993; Ichiye & Karplus, 1991). In the case of a GPCR, for instance, the relevant functional motions might involve the outward swing motion of TM6 due to activation, or differential rearrangements induced by biased ligands and connected to functional selectivity (Whalen, Rajagopal, & Lefkowitz, 2011). In the case of

arrestin, the activation transition involves the rotation of the C-domain relative to the N-domain and the exposure of the C-tail (Shukla et al., 2013). It is therefore interesting to extract such collective motions from the overall dynamics. The principal component analysis of the covariance matrix decomposes by diagonalization the covariance matrix (see Eq. 9) into collective modes at different frequencies. The low frequency eigenvectors are the principal modes connected to longer time scale collective motion. The dynamics along the individual modes can be inspected and visualized separately, thereby allowing one to filter the main modes of collective motion from local fluctuations. The former defines the so-called "Essential Dynamics," related to the idea that these are the modes essential for function.

Different from the other popular analysis tool called NMA, the Essential Dynamics is not based on a harmonic approximation of the potential energy and this allows one to study anharmonic features in protein motion. This reflects the notion that at physiological temperatures, protein dynamics on longer timescales can be described as diffusion between local minima, with anharmonic character, whereas only on short timescales it is given by fluctuations within a single minimum.

In practice, the first step is the calculation of the $3N \times 3N$ covariance matrix C_{ij} of the protein, after aligning the trajectory to a reference structure:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{9}$$

where the brackets indicate as usual the time average over the production part of the trajectory (approximating the ensemble average) and x_i is one of the three spatial coordinates of an atom, while x_j is one of the three coordinates of another atom. The trajectory can also be made of crystal structures.

Upon diagonalization of the covariance matrix, $3N$ eigenvectors and eigenvalues are retrieved. Six of them are approximately zero and represent the global rotation and translation degrees of freedom, the remaining $3N-6$ can be put in decreasing order, so that the first ones on the list are the highest eigenvalues that cover the most significant percentage of the variance.

In GROMACS, the covariance matrix is calculated and diagonalized with the single command g_covar, and it is possible to align the trajectory and calculate the covariance of distinct subgroups. Typically, in order to be able to identify meaningful global motions while keeping the dimension under control, a good choice is to calculate the covariance matrix on the Cα atoms. Several options are available for the output. Besides eigenvalues and eigenvectors, one can calculate the reduced covariance matrix, where the three coordinates of each degree of freedom are combined to a single value for visualization purposes. The eigenvectors and eigenvalues can be further analyzed with the program g_anaeig, which allows one to project the trajectory on the eigenvector (option –proj), to visualize and evaluate the kind of collective motion by printing the .pdb file of the projection or the extreme structures along the projection (–extr), and to calculate the amount of fluctuations pertaining to a given mode (see manual). When comparing the Essential Dynamics analysis of different trajectories of the same system or of separate parts of a single trajectory, it is important to assess convergence: to this end, the superposition of the subspaces generated by the first few (e.g., Eq. 10) eigenvectors should be compared. The similarity of the

subspaces, can be evaluated through the RMSIP measure or some other comparison criterion (Zen et al., 2008):

$$RMSIP = \sqrt{\sum_{i,j=1}^{10} \frac{\left|\vec{v_i} \cdot \vec{w_j}\right|}{10}} \tag{10}$$

where $\left\{\vec{v_i}\right\}$ is the first set of eigenvectors and $\left\{\vec{w_j}\right\}$ is the second set. The useful package of R scripts that can be used to perform these and other analyses is Bio3d (http://thegrantlab.org/bio3d/index.php).

A practical application of the covariance analysis is the comparison between different single point mutants or different ligand bound states. In such cases, the different trajectories may be concatenated into one metatrajectory on which the PCA is carried out. The individual trajectories can be then separately projected onto the set of modes, allowing a direct comparison of the collective motions of each ligand state. As an example, in the case of visual Arrestin, the PCA analysis of the concatenated metatrajectory of wild-type protein and R175E mutant shows that the first eigenvector (Fig. 4A) recapitulates the activation transition observed in

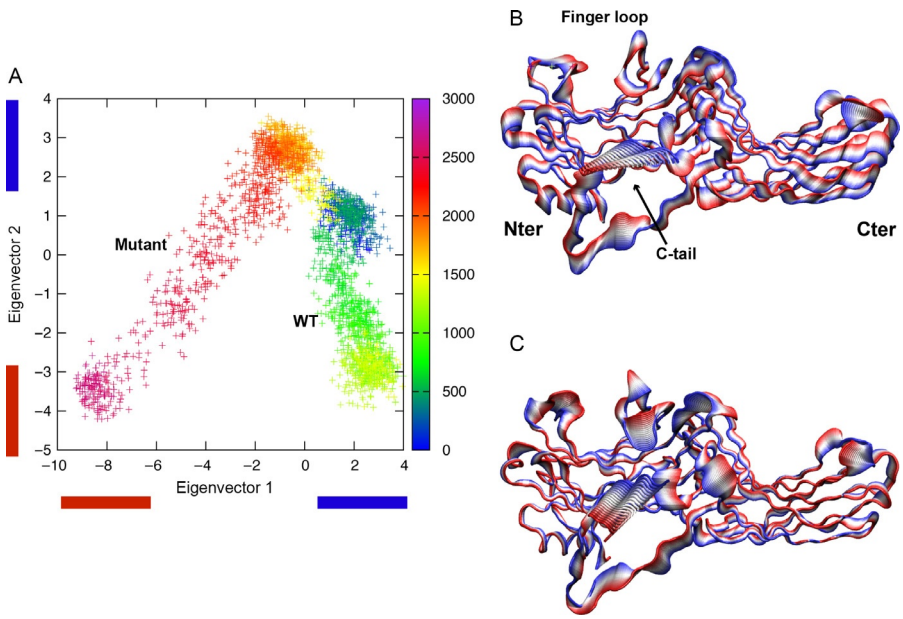

FIG. 4

Essential Dynamics analysis of visual Arrestin obtained concatenating a MD trajectory of the WT protein and one of the constitutively active R175E mutant (Sensoy et al., 2016). (A) 2D projection on eigenvectors 1 and 2 which shows the motion along 1 of the mutant and not of the WT. Color bar reporting time evolution. (B) Structural representation of motion along eigenvector 1 and (C) same as B on eigenvector 2. The *blue* and *red colors* refer to the extreme values of the projection in (A). For details on the MD simulation, see Sensoy et al. (2016).

the mutant R175E, including the C-domain rigid motion and the displacement of the C-tail, see Fig. 4B and Sensoy et al. (2016). In contrast, eigenvector 2 describes motions occurring in both systems and involving the fluctuation of the finger loop and other loop regions (Fig. 4C).

ACKNOWLEDGMENTS

I.S.M. acknowledges support by the Fundação para a Ciência e a Tecnologia (FCT) Investigator programme—IF/00578/2014 (cofinanced by European Social Fund and Programa Operacional Potencial Humano) and a Marie Skłodowska-Curie Individual Fellowship MSCA-IF-2015 [MEMBRANEPROT 659826]. This work was also financed 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 Internationalization and Portuguese national funds via FCT, under project POCI-01-0145-FEDER-007440.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Amadei, A., Linssen, A. B. M., & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins*, *17*, 412–425.

Arango-Lievano, M., Sensoy, O., Borie, A., Corbani, M., Guillon, G., Sokoloff, P., et al. (2016). A GIPC1-palmitate switch modulates dopamine Drd3 receptor trafficking and signaling. *Molecular and Cellular Biology*, *36*, 1019–1031.

Ballesteros, J. A., & Weinstein, H. (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G protein-coupled receptors. In C. S. Stuart (Ed.), *Methods in neurosciences: Vol. 25* (pp. 366–428): San Diego, CA: Academic Press.

Benner, S. A., & Gerloff, D. (1991). Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: A prediction of the structure of the catalytic domain of protein kinases. *Advances in Enzyme Regulation*, *31*, 121–181.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Di Nola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*, 3684–3690.

Berendsen, H. J. C., Spoel, D. V. D., & Drunen, R. V. (1995). GROMACS: A message passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*, 43–56.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(Web Server issue), W252–258. https://doi.org/10.1093/nar/gku340.

Brooks, B., & Karplus, M. (1985). Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(15), 4995–4999.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins*, *21*(3), 167–195.

Bustamante, J. G. F. R. C. (2003). Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(22), 12564–12569.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

Carpenter, E. P., Beis, K., Cameron, A. D., & Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology*, *18*(5), 581–586. https://doi.org/10.1016/j.sbi.2008.07.001.

Carpenter, B., Nehme, R., Warne, T., Leslie, A. G., & Tate, C. G. (2016). Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature*, *536*(7614), 104–107. https://doi.org/10.1038/nature18966.

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., et al. (2013). ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, *53*(3–4), 199–206. https://doi.org/10.1002/ijch.201200096.

Chiappori, F., Merelli, I., Colombo, G., Milanesi, L., & Morra, G. (2012). Molecular mechanism of allosteric communication in Hsp70 revealed by molecular dynamics simulations. *PLoS Computational Biology, 8*, e1002844 Figshare 1.

Chodera, J. D. (2016). A simple method for automated equilibration detection in molecular simulations. *Journal of Chemical Theory and Computation*, *12*(4), 1799–1805. https://doi.org/10.1021/acs.jctc.5b00784.

Choi, Y. S., Yang, J. S., Choi, Y., Ryu, S. H., & Kim, S. (2009). Evolutionary conservation in multiple faces of protein interaction. *Proteins*, *77*, 14–25.

Chung, D. A., Zuiderweg, E. R., Fowler, C. B., Soyer, O. S., Mosberg, H. I., & Neubig, R. R. (2002). NMR structure of the second intracellular loop of the alpha 2A adrenergic receptor: Evidence for a novel cytoplasmic helix. *Biochemistry*, *41*(11), 3596–3604.

Colizzi, F., Perozzo, R., Scapozza, L., Recanatini, M., & Cavalli, A. (2010). Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *Journal of the American Chemical Society*, *132*, 7361–7371.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, R. I., Merz, K. M. J., Ferguson, D. M., et al. (1995). The Amber force field. *Journal of the American Chemical Society*, *117*, 5179–5197.

Costanzi, S. (2012). Homology modeling of class a G protein-coupled receptors. *Methods in Molecular Biology*, *857*, 259–279. https://doi.org/10.1007/978-1-61779-588-6_11.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*, 10089–10092.

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., & Mark, A. E. (1999). Peptide folding: When simulation meets experiment. *Angewandte Chemie, International Edition*, *38*, 236–240.

de Vries, S. J., & Bonvin, A. M. J. J. (2011). CPORT: A consensus Interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS One, 6*(3): e17695.

de Vries, S. J., van Dijk, M., & Bonvin, A. M. J. J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, *5*(5), 883–897. https://doi.org/10.1038/nprot.2010.32.

Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A protein–protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, *125*(7), 1731–1737. https://doi.org/10.1021/ja026939x.

Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., et al. (2011). Activation mechanism of the β2-adrenergic receptor. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(46), 18684–18689.

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113. https://doi.org/10.1186/1471-2105-5-113.

Esguerra, M., Siretskiy, A., Bello, X., Sallander, J., & Gutierrez-de-Teran, H. (2016). GPCR-ModSim: A comprehensive web based solution for modeling G-protein coupled receptors. *Nucleic Acids Research*, *44*(W1), W455–462. https://doi.org/10.1093/nar/gkw403.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., et al. (2006). Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics* . 0 5, Unit-5.6. https://doi.org/10.1002/0471250953.bi0506s15.

Feller, S. E., & Pastor, R. W. (1996). On simulating lipid bilayers with an applied surface tension: Periodic boundary conditions and undulations. *Biophysical Journal*, *71*(3), 1350–1355.

Ghosh, E., Kumari, P., Jaiman, D., & Shukla, A. K. (2015). Methodological advances: The unsung heroes of the GPCR structural revolution. *Nature Reviews. Molecular Cell Biology*, *16*(2), 69–81. https://doi.org/10.1038/nrm3933.

Giorgino, T., & De Fabritiis, G. (2011). High-throughput steered molecular dynamics study on the free energy profile of ion permeation through gramicidin A. *Journal of Chemical Theory and Computation*, *7*, 1943–1950.

Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., & Caves, L. S. (2006). Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, *22*(21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461.

Grubmuller, H., Heymann, B., & Tavan, P. (1996). Ligand binding: Molecular mechanics calculation of the streptavidin–biotin rupture force. *Science*, *271*, 997–999.

Guharoy, M., & Chakrabarti, P. (2005). Conservation and relative importance of residues across protein–protein interfaces. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15447–15452. https://doi.org/10.1073/pnas.0505425102.

Guixa-Gonzalez, R., Ramırez-Anguita, J., Kaczor, A. A., & Selent, J. (2012). Simulating G-protein coupled receptors in native-like membranes: From monomer to oligomers. *Methods in Cell Biology*, *117*, 63–90.

Gwan, J. F., & Baumgaertner, A. (2007). Cooperative transport in a potassium ion channel. *The Journal of Chemical Physics*, *127*, 045103.

Han, Y., Moreira, I. S., Urizar, E., Weinstein, H., & Javitch, J. A. (2009). Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nature Chemical Biology*, *5*(9), 688–695. https://doi.org/10.1038/nchembio.199.

Hansen, D. F., Vallurupalli, P., & Kay, L. E. (2008). Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *Journal of Biomolecular NMR*, *41*(3), 113. https://doi.org/10.1007/s10858-008-9251-5.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *28*(1), 100–108. https://doi.org/10.2307/2346830.

Hatrick, K., & Taylor, W. R. (1994). Sequence conservation and correlation measures in protein structure prediction. *Computers & Chemistry*, *18*(3), 245–249.

Hayat, S., Sander, C., Marks, D. S., & Elofsson, A. (2015). All-atom 3D structure prediction of transmembrane beta-barrel proteins from sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5413–5418. https://doi.org/10.1073/pnas.1419956112.

Hermans, E. (2003). Biochemical and pharmacological control of the multiplicity of coupling at G-protein-coupled receptors. *Pharmacology & Therapeutics*, *99*(1), 25–44.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447.

Hopf, T. A., Schärfe, C. P., Rodrigues, J. P., Green, A. G., Kohlbacher, O., Sander, C., et al. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife*, *3*, e03430.

Hu, Z., Ma, B., Wolfson, H., & Nussinov, R. (2000). Conservation of polar residues as hot spots at protein interfaces. *Proteins*, *39*(4), 331–342.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. 27-38.

Ichiye, T., & Karplus, M. (1991). Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins*, *11*(3), 205–217.

Israelewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., & Schulten, K. (2001). Steered molecular dynamics investigations of protein function. *Journal of Molecular Graphics & Modelling*, *19*, 13–25.

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., et al. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins*, *55*(2), 351–367. https://doi.org/10.1002/prot.10613.

Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, *78*(14), 2690–2693.

Jia, L., Chisari, M., Maktabi, M. H., Sobieski, C., Zhou, H., Konopko, A. M., et al. (2014). A mechanism regulating GPCR signaling that requires cycles of protein palmitoylation and depalmitoylation. *The Journal of Biological Chemistry*, *289*, 6249–6257.

Jose Manuel Perez-Aguilar, J. S., LeVine, M. V., Khelashvili, G., & Weinstein, H. (2014). A functional selectivity mechanism at the serotonin-2A GPCR involves ligand-dependent conformations of intracellular loop 2. *Journal of the American Chemical Society*, *136*(45), 16044–16054.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*, 2577–2637.

Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., et al. (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature*, *523*(7562), 561–567. https://doi.org/10.1038/nature14656.

Khelashvili, G., Grossfield, A., Feller, S. E., Pitman, M. C., & Weinstein, H. (2009). Structural and dynamic effects of cholesterol at preferred sites of interaction with rhodopsin identified from microsecond length molecular dynamics simulations. *Proteins*, *76*, 413–417.

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, *32*(Web Server issue), W526–531. https://doi.org/10.1093/nar/gkh468.

Kobilka, B., & Schertler, G. F. (2008). New G-protein-coupled receptor crystal structures: Insights and limitations. *Trends in Pharmacological Sciences*, *29*(2), 79–83. https://doi.org/10.1016/j.tips.2007.11.009.

Lanctot, P. M., Leclerc, P. C., Clement, M., Auger-Messier, M., Escher, E., Leduc, R., et al. (2005). Importance of N-glycosylation positioning for cell-surface expression, targeting, affinity and quality control of the human AT1 receptor. *The Biochemical Journal*, *390*(Pt. 1), 367–376. https://doi.org/10.1042/bj20050189.

Levitt, M., Sander, C., & Stern, P. S. (1985). Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology*, *181*(3), 423–447.

Li, W., McWilliam, H., Goujon, M., Cowley, A., Lopez, R., & Pearson, W. R. (2012). PSI-search: Iterative HOE-reduced profile SSEARCH searching. *Bioinformatics*, *28*(12), 1650–1651. https://doi.org/10.1093/bioinformatics/bts240.

Lu, H., Isralewitz, B., Krammer, A., Vogel, V., & Schulten, K. (1998). Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, *75*, 662–671.

Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, *13*(3), 373–380. https://doi.org/10.1016/j.str.2005.02.002.

Mackerell, A. D., Feig, M., & Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, *25*(11), 1400–1415. https://doi.org/10.1002/jcc.20065.

Maiorov, V. N., & Crippen, G. M. (1995). Size independent comparison od protein 3-dimensional structures. *Proteins*, *22*, 273–283.

Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., et al. (2011). Protein 3D structure computed from evolutionary sequence variation. *PLoS One, 6*(12). e28766https://doi.org/10.1371/journal.pone.0028766.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & de Vries, A. H. (2007). The Martini force field: Coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry. B*, *111*, 7812–7824.

Martinez-Archundia, M., Cordomi, A., Garriga, P., & Perez, J. J. (2012). Molecular modeling of the M3 acetylcholine muscarinic receptor and its binding site. *Journal of Biomedicine & Biotechnology*, *2012*, 789741. https://doi.org/10.1155/2012/789741.

Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987). Interior and surface of monomeric proteins. *Journal of Molecular Biology*, *196*(3), 641–656.

Miller, S., Lesk, A. M., Janin, J., & Chothia, C. (1987). The accessible surface area and stability of oligomeric proteins. *Nature*, *328*(6133), 834–836. https://doi.org/10.1038/328834a0.

Monticelli, L., & Tieleman, D. P. (2013). Force fields for classical molecular dynamics. In *Biomolecular simulations* (pp. 197–213). Springer.

Moreira, I. S. (2014). Structural features of the G-protein/GPCR interactions. *Biochimica et Biophysica Acta*, *1840*(1), 16–33. https://doi.org/10.1016/j.bbagen.2013.08.027.

Moreira, I. S. (2015). The role of water occlusion for the definition of a protein binding hotspot. *Current Topics in Medicinal Chemistry*, *15*(20), 2068–2079.

Moreira, I. S., Shi, L., Freyberg, Z., Ericksen, S. S., Weinstein, H., & Javitch, J. A. (2010). Structural basis of dopamine receptor activation. In K. A. Neve (Ed.), *The dopamine receptors* (pp. 47–73). Totowa, NJ: Humana Press.

Nagle, J. F., & Tristram-Nagle, S. (2000). Structure of lipid bilayers. *Biochimica et Biophysica Acta*, *1469*(3), 159–195.

Niv, M. Y., & Filizola, M. (2008). Influence of oligomerization on the dynamics of G-protein coupled receptors as assessed by normal mode analysis. *Proteins*, *71*(2), 575–586. https://doi.org/10.1002/prot.21787.

Okada, T., Sugihara, M., Bondar, A. N., Elstner, M., Entel, P., & Buss, V. (2004). The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. *Journal of Molecular Biology*, *342*(2), 571–583. https://doi.org/10.1016/j.jmb.2004.07.044.

Okazaki, K., Koga, N., Takada, S., Onuchic, J. N., & Wolynes, P. G. (2006). Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 11844–11849.

Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M., & Jensen, J. H. (2011). PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation*, *7*(2), 525–537. https://doi.org/10.1021/ct100578z.

Patra, S. M., Chakraborty, S., Shahane, G., Prasanna, X., Sengupta, D., Maiti, P. K., et al. (2015). Differential dynamics of the serotonin1a receptor in membrane bilayers of varying cholesterol content revealed by all atom molecular dynamics simulation. *Molecular Membrane Biology*, *32*, 127–137.

Pearson, W. R. (2014). BLAST and FASTA similarity searching for multiple sequence alignment. *Methods in Molecular Biology*, *1079*, 75–101. https://doi.org/10.1007/978-1-62703-646-7_5.

Periole, X., Cavalli, M., Marrink, S. J., & Ceruso, M. (2009). Combining an elastic network with a coarse-grained molecular force field: Structure, dynamics, and intermolecular recognition. *Journal of Chemical Theory and Computation*, *5*, 2531–2543.

Petrone, P., & Pande, V. S. (2006). Can conformational change be described by only a few normal modes? *Biophysical Journal*, *90*(5), 1583–1593. https://doi.org/10.1529/biophysj.105.070045.

Probst, W. C., Snyder, L., Schuster, D. I., Brosius, J., & Sealfon, S. C. (1992). Sequence alignment of the G-protein coupled receptor superfamily. *DNA and Cell Biology*, *11*, 1–20.

Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., & Ben-Tal, N. (2002). Rate4Site: An algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, *18*, S71–7.

Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., et al. (2011). Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature*, *469*(7329), 175–180. https://doi.org/10.1038/nature09648.

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, *344*(6191), 1492–1496. https://doi.org/10.1126/science.1242072.

Rzepiela, A. J., Schäfer, L. V., Goga, N., Risselada, H. J., De Vries, A. H., & Marrink, S. J. (2010). Reconstruction of atomistic details from coarse-grained structures. *Journal of Computational Chemistry*, *31*, 1333–1343.

Samuel Hertig, N. R. L., & Dror, R. O. (2016). Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. *PLoS Computational Biology, 12*(6), e1004746.

Schrödinger, L. L. C. (2015). *The PyMOL Molecular Graphics System, Version 1.8.*

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, *33*(Web Server issue), W382–388. https://doi.org/10.1093/nar/gki387.

Seelig, A., & Seelig, J. (1974). The dynamic structure of fatty acyl chains in a phospholipid bilayer measured by deuterium magnetic resonance. *Biochemistry*, *13*(23), 4839–4845.

Sensoy, O., Atılgan, A. R., & Atılgan, C. (2017). FbpA iron storage and release are governed by periplasmic microenvironments. *Physical Chemistry Chemical Physics*, *19*, 6064–6075.

Sensoy, O., Moreira, I. S., & Morra, G. (2016). Understanding the differential selectivity of arrestins toward the phosphorylation state of the receptor. *ACS Chemical Neuroscience*, *7*(9), 1212–1224. https://doi.org/10.1021/acschemneuro.6b00073.

Sensoy, O., & Weinstein, H. (2015). A mechanistic role of helix 8 in GPCRs: Computational modeling of the dopamine D2 receptor interaction with the GIPC1–PDZ-domain. *Biochimica et Biophysica Acta Biomembranes*, *1848*(4), 976–983.

Shukla, A. K., Manglik, A., Kruse, A. C., Xiao, K., Reis, R. I., Tseng, W.-C., et al. (2013). Structure of active b-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature*, *497*, 137–141.

Sievers, F., & Higgins, D. G. (2014). Clustal omega, accurate alignment of very large numbers of sequences. *Methods in Molecular Biology*, *1079*, 105–116. https://doi.org/10.1007/978-1-62703-646-7_6.

Soding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*(Web Server issue), W244–248. https://doi.org/10.1093/nar/gki408.

Tai, K., Fowler, P., Mokrab, Y., Stansfeld, P., & Sansom, M. S. (2008). Molecular modeling and simulation studies of ion channel structures, dynamics and mechanisms. *Methods in Cell Biology*, *90*, 233–265.

Team, R. D. C. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., et al. (2006). Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie, International Edition*, *45*(25), 4064–4092. https://doi.org/10.1002/anie.200502655.

Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., et al. (2016). The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, *428*(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.

Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., & Oliva, R. (2011). COCOMAPS: A web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, *27*(20), 2915–2916. https://doi.org/10.1093/bioinformatics/btr484.

Webb, B., & Sali, A. (2014). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, *47*, 5.6.1–5.6.32. https://doi.org/10.1002/0471250953.bi0506s47.

Whalen, E. J., Rajagopal, S., & Lefkowitz, R. J. (2011). Therapeutic potential of beta-arrestin- and G protein-biased agonists. *Trends in Molecular Medicine*, *17*(3), 126–139. https://doi.org/10.1016/j.molmed.2010.11.004.

Wolynes, P. G., Onuchic, J. N., & Thirumalai, D. (1995). Navigating the folding routes. *Science*, *267*(5204), 1619–1620. https://doi.org/10.1126/science.7886447.

Worth, C. L., Kleinau, G., & Krause, G. (2009). Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS One:4*(9). e7011https://doi.org/10.1371/journal.pone.0007011.

Worth, C. L., Kreuchwig, A., Kleinau, G., & Krause, G. (2011). GPCR-SSFE: A comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics*, *12*, 185. https://doi.org/10.1186/1471-2105-12-185.

Xu, W., Harrison, S. C., & Eck, M. J. (1997). Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, *385*, 595–602.

Zen, A., Carnevale, V., Lesk, A. M., & Micheletti, C. (2008). Correspondences between low-energy modes in enzymes: Dynamics-based alignment of enzymatic functional families. *Protein Science*, *17*, 918–929.

Zheng, H., Loh, H., & Law, P. Y. (2013). Posttranslation modification of G protein-coupled receptor in relationship to biased agonism. *Methods in Enzymology*, *522*, 391–408.

Zvelebil, M. J., Barton, G. J., Taylor, W. R., & Sternberg, M. J. (1987). Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *Journal of Molecular Biology*, *195*(4), 957–961.