

# Hybrid Molecular Mechanics/Coarse-Grained Simulations for Structural Prediction of G-Protein Coupled Receptor/Ligand Complexes

Michael Leguèbe<sup>1,2,3</sup>, Chuong Nguyen<sup>2,3,9</sup>, Luciana Capece<sup>2,3,4,9</sup>, Zung Hoang<sup>5</sup>, Alejandro Giorgetti<sup>2,3,6\*</sup>, Paolo Carloni<sup>2,3,7</sup>

**1** INRIA Bordeaux Sud-Ouest, Institut de Mathématiques de Bordeaux, Talence, France, **2** Computational Biophysics, German Research School for Simulation Sciences, Jülich, Germany, **3** JARA - High-Performance Computing, Jülich, Germany, **4** International Centre for Genetic Engineering and Biotechnology, Trieste, Italy, **5** Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam, **6** Department of Biotechnology, University of Verona, Verona, Italy, **7** Institute for Advanced Simulation, IAS-5 Institute for Computational Biomedicine, Forschungszentrum Jülich, Jülich, Germany

## Abstract

Understanding how ligands bind to G-protein coupled receptors (GPCRs) provides insights into a myriad of cell processes and is crucial for drug development. Here we extend a hybrid molecular mechanics/coarse-grained (MM/CG) approach applied previously to enzymes to GPCR/ligand complexes. The accuracy of this method for structural predictions is established by comparison with recent atomistic molecular dynamics simulations on the human  $\beta_2$  adrenergic receptor, a member of the GPCRs superfamily. The results obtained with the MM/CG methodology show a good agreement with previous all-atom classical dynamics simulations, in particular in the structural description of the ligand binding site. This approach could be used for high-throughput predictions of ligand poses in a variety of GPCRs.

**Citation:** Leguèbe M, Nguyen C, Capece L, Hoang Z, Giorgetti A, et al. (2012) Hybrid Molecular Mechanics/Coarse-Grained Simulations for Structural Prediction of G-Protein Coupled Receptor/Ligand Complexes. PLoS ONE 7(10): e47332. doi:10.1371/journal.pone.0047332

**Editor:** Jörg Langowski, German Cancer Research Center, Germany

**Received:** May 16, 2012; **Accepted:** September 14, 2012; **Published:** October 19, 2012

**Copyright:** © 2012 Leguèbe et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work has been supported in part by a grant from the Deutsche Forschungsgemeinschaft (CA 973/6-1), by a grant from Vietnam National Foundation for Science and Technology Development (DFG.2011.01), and by a grant from the Project ISS UPR- 20009-1301355 (CUP I85J08000040005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: alejandro.giorgetti@univr.it

<sup>9</sup> These authors contributed equally to this work.

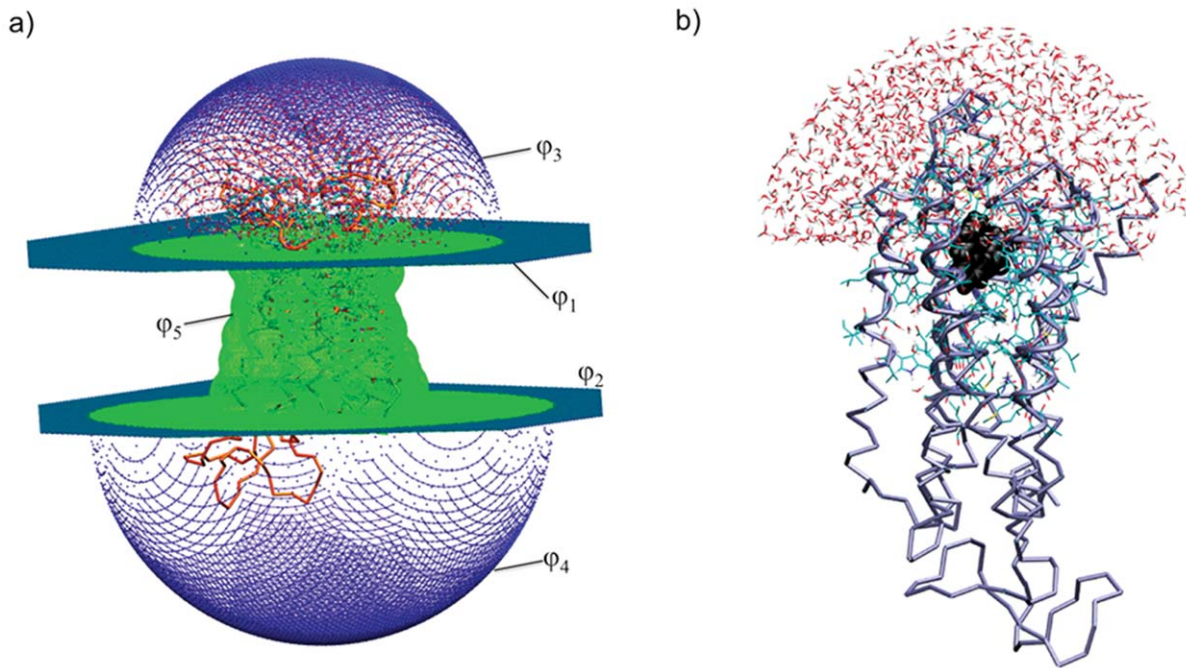
## Introduction

G-protein coupled receptors (GPCRs) are involved in an enormous number of biochemical processes at the cell membrane. They comprise the largest membrane protein superfamily across eukaryotes [1]. Human GPCRs are among the most important targets of pharmaceutical intervention, constituting the target for ~30% of clinically used drugs [2]. Thus, methods for investigating how ligands bind to GPCRs are crucial not only for characterizing processes in cells but also for drug development.

Experimental structures of GPCRs are available for eleven members of that superfamily in eukaryotes (Table S1). Molecular dynamics (MD) simulations, based on structures predicted by bioinformatics methods, are often used to identify ligand poses on GPCRs for which experimental structural information is not available [3]. This approach can be very CPU intensive [4], especially to characterize large numbers of ligand/receptor complexes. On the other hand, coarse-grained (CG) approaches allow the study of large systems on longer timescales than those usually explored with MD [5,6]. Indeed, the reduction of the number of degrees of freedom allows a reduction of the simulation time by ~2 to ~3 orders of magnitude compared to full atom force fields [7]. However, these approaches cannot describe the intermolecular ligand/protein interactions at atomic detail as required in ligand pose predictions.

A possible strategy to address this issue is to combine atomistic with CG modeling [8–14]. Molecular Mechanics/Coarse-Grained (MM/CG) simulation is an approach in which different representations of the system are modeled concurrently (Figure 1). A coupling scheme is used to connect the boundary of models. This approach has been developed for proteins by several groups, including ours [10–14]. In our scheme [12,13], a region of interest (i.e. the active site of an enzyme, MM region) is treated at a molecular level using an atomistic force field. The one used here is the GROMOS96 43a1 force field [15,16]. Hydration is accounted for including a droplet of water molecules around the MM region. The protein frame is described at a CG level using a Go-like model [17]. Such a model includes only the backbone  $C\alpha$  atoms. An interface (I region) is defined between the MM and CG regions to bridge the two different resolution models.

Our MM/CG approach was originally developed for enzymes where the MM region is exposed to the water solvent [13]. In GPCRs, the binding sites are located in the transmembrane region [18]. Thus, the presence of the lipid bilayer should be taken into account. In addition, the model requires some modifications to avoid the diffusion of water into the hydrophobic regions of the lipid bilayer. Hence, we have modified our MM/CG scheme and tested the accuracy of this method by comparing it with recent extensive atomistic molecular dynamics simulations [19] on the



**Figure 1. Simulated system.** a) Five walls around the GPCR are used to mimic the presence of the lipid bilayer: the planar walls ( $\varphi_{1,2}$ ) are the dark green sheets located at the height of the membrane lipids head, the outer walls ( $\varphi_{3,4}$ ) are the dark blue hemispheres, and the membrane wall ( $\varphi_5$ ) is the surface in green. b) Model of h $\beta$ 2-AR in complex with S-Carazolol (S-Car) for the MM/CG simulations; the protein backbone is represented by tubes, the MM and I regions and the water molecules are represented in licorice form, and the ligand S-Car is represented with black spheres. doi:10.1371/journal.pone.0047332.g001

human  $\beta$ 2 adrenergic receptor (h $\beta$ 2-AR), a member of the GPCRs superfamily involved in cardiac functions [20,21].

## Methods

### The MM/CG Scheme

The total potential energy of the system in the MM/CG approach is split into terms corresponding to different sets of atoms, belonging to the MM, CG and CG/I regions (see [12]) as follows (Eq. 1):

$$V = E_{MM} + E_I + E_{CG} + E_{MM/I} + E_{CG/I} + E_{SD} \quad (1)$$

Here, the first term denotes the GROMOS96 force field [15]. Atoms in the interface region are represented at the atomistic level, thus the terms  $E_I$  and  $E_{MM/I}$  have the same form as  $E_{MM}$ .  $E_{CG}$  is given by a Go-like model [22] (Eq. 2):

$$E_{CG} = \frac{1}{4} \sum_i K_b (|R_i - R_{i+1}|^2 - b_{ii+1}^2)^2 + \sum_{i>j} V_0 \{1 - \exp[-B_{ij}(|R_i - R_j| - b_{ij})]\}^2 \quad (2)$$

(i) The first term describes the interaction between consecutive CG beads (the C $\alpha$  atoms);  $K_b$  is the force constant;  $b_{ij}$  is the equilibrium distance, corresponding to the native distance between CG atoms; and (ii) Non-bonded interactions are taken into account in the second (Morse-potential-type) term:  $V_0 = 5.3$  kJ mol $^{-1}$  is the well depth, and  $B_{ij}$  is its modulating coefficient which is a function of  $b_{ij}$ :  $B_{ij} = 6/b_{ij}$  (nm $^{-1}$ ). These two parameters have

been already used for both soluble and membrane enzymes [12,13]. Here we consider the same value for  $V_0$  while  $B_{ij}$  is set to  $5+6/b_{ij}$  (nm $^{-1}$ ). This setup ensures the stability of the protein inside its transmembrane site. The interaction between the CG and I regions, described by the term  $E_{CG/I}$ , is treated in the same form as  $E_{CG}$ . The bonded terms are included between CG atoms and consecutive C $\alpha$  atoms of I region, and the non-bonded interactions between CG atoms and both C $\alpha$  and C $\beta$  atoms [12]. This interface potential ensures the integrity of the backbone. The  $E_{SD}$  term mimics the thermal and viscous solvent effects acting on the system [23].

To take into account the presence of the lipid bilayer, we have modified the potential energy function as follows.

- 1) We introduced five walls around the protein (Figure 1a). These are described by five functions  $\varphi_i$  ( $i=1-5$ ) using a level-set approach [24]. The region where the set of  $\varphi_i$  is positive characterizes the protein site. The wall itself is the set of points for which the set  $\varphi_i$  vanishes (Figure S1b). Two planar walls ( $\varphi_i$ ,  $i=1,2$ ) coincide with the height of the membrane lipid's head. Two hemispheric walls ('outer walls',  $\varphi_i$ ,  $i=3,4$ ) capping the extracellular and cytoplasmic ends of the protein (Figure 1a) are defined as follow:

$$\varphi_i(\mathbf{r}) = r_i - \|\mathbf{r} - \mathbf{c}_{hi}\|, \quad i=3,4 \quad (3)$$

where their locations are defined only outside the membrane region. The center  $\mathbf{c}_i$  of each hemisphere is located at the height of the membrane lipid heads, above/under the center of mass of the protein. The radius  $r_i$  of each hemisphere is such that the minimum distance between any protein atoms and the wall is 20 Å. This creates a droplet of waters around the MM region similar to ref. [13]. The last wall ('membrane

wall',  $\phi_5$ ) follows the initial shape of the interface between protein and membrane (Figure 1a). This wall is defined as:

$$\phi_5(\mathbf{r}) = r_p - \min_j \|\mathbf{r} - \mathbf{c}_j\| \quad (4)$$

where the distance between the point  $\mathbf{r}$  and the closest initial position of C $\alpha$  atoms  $\mathbf{c}_j$  is computed, and  $r_p$  is a distance parameter that we have set to 2.0 Å. Additionally a smoothing technique (see Text S1 for further details) is applied to avoid the discontinuities in the wall.

- 2) Boundary potentials are added to the MM/CG potential energy function. They are defined as functions of the distance ( $V_i(d)$  ( $i = 1-5$ ),  $d = \min(\phi_i$ ,  $i = 1-5$ ) from the corresponding walls, as follows: (Figure S1)

$$V_i(d) = 1/d, i = 1, 2 \quad (5)$$

$$V_i(d) = 4\epsilon \left[ \left( \frac{\sigma}{d} \right)^2 - \left( \frac{\sigma}{d} \right) \right], i = 3, 4, 5 \quad (6)$$

The index  $i$  of boundary potential corresponds to the index of the surfaces  $\phi_i$ . The potential applied to an atom is the one corresponding to the closest wall from that atom  $d = \min \phi_i$  ( $i = 1, 5$ ).  $V_i$  ( $i = 1, 2$ ) is purely repulsive;  $V_i$  ( $i = 3, 4, 5$ ) is a softened Lennard-Jones-type potential;  $\epsilon$  is the depth of the potential well; and  $\sigma$  is the finite distance at which the potential  $V_i$  ( $i = 3, 4, 5$ ) is zero. The minimum of the potential is at  $d = r_p$ . Waters,  $\alpha$ -carbons of both MM and CG regions, and atoms belonging to external aromatic residues TRP and TYR are influenced by these potentials. The potential on the membrane wall is represented by a softened Lennard-Jones potential,  $V_5$ . It constrains the shape of the protein while providing a good degree of flexibility. This model neither includes electrostatics nor allows distinguishing between different types of bilayers.

The force due to the presence of the wall is derived from the potentials  $V_i$ :

$$\vec{F}_i(\mathbf{r}) = - \frac{\partial V_i}{\partial d} \nabla d(\mathbf{r}) \quad (7)$$

The cutoff distance of the force is set to 7 Å for the repelling walls  $V_i$  ( $i = 1, 2$ ), and to  $1.5r_p$  for the outer walls and membrane wall  $V_i$  ( $i = 3, 4, 5$ ). The first value is chosen such that a water molecule can not pass through this distance during one time step, while the second value guarantees that the force does not affect the MM region. The force is shifted so that it is continuous at the cutoff distance, to avoid a sharp disruption. In addition, it is set to a finite value ( $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ ) near the wall to prevent too large of a force being applied to the system.

### MM/CG Simulations of Human $\beta_2$ -adrenergic Receptor (h $\beta_2$ AR) in Complex with S-carazolol (S-Car) or R-isoprenaline (R-Iso)

The calculations are based on the X-ray structure of the h $\beta_2$ AR in complex with S-Car (PDB code 2RH1) [25]. Since the third intracellular loop (ICL3, residues 231–262) is not present in the structure, it was predicted using the Modeller9v3 program [26]. The structure of the complex between the h $\beta_2$ AR and the agonist R-Iso was obtained following the procedure of [19]. 975 SPC

water molecules [27] were added. This constitutes a layer of approximately  $\sim 15$  Å around the MM region. Similarly to our previous works [12,13], the MM region included all the residues at a distance less than 5 Å from the bound ligand, which consisted in residues 79–82, 86, 109 to 118, 164–165, 193–195, 199–208, 282, 286, 289–290, 293, 308, 311–316, the corresponding ligands R-Iso or S-Car, and the water molecules. A cutoff of 6 Å (as measured from the MM boundary) was applied in order to calculate the residues included in the interface between the MM and the CG subsystems. The MM and I regions were treated using the GROMOS96 43a1 force field [15,16]. Our MM/CG scheme has been extensively tested for this force field [12,13]. Structural properties for a large number of proteins do not change by a large amount when using updated versions of this force field, such as 53a5 and 53a6 [16]. The resulting total number of atoms is 4597 and 4587 for the h $\beta_2$ AR.S-Car and h $\beta_2$ AR.R-ISO complexes respectively. Starting from this structure, 800 ns MM/CG simulations were performed using a 2 fs time step. The protein complexes were encapsulated in a  $\sim 31$  Å thick implicit membrane, with the transmembrane wall 2.0 Å from the C $\alpha$  atoms. Cutoffs of 16 Å were used for the electrostatic, van der Waals and Go-like interactions. The SHAKE algorithm [28] was used to fix the distance in bonds containing hydrogen(s). The temperature was set to 300 K using stochastic dynamics, controlled by an inverse friction constant with a value of 0.4 ps. Periodic boundary conditions were used. RESP charges [29,30] for the ligands were taken from ref. [19]. All simulations have been performed using our MM/CG implementation in Gromacs 4.5.1 [31].

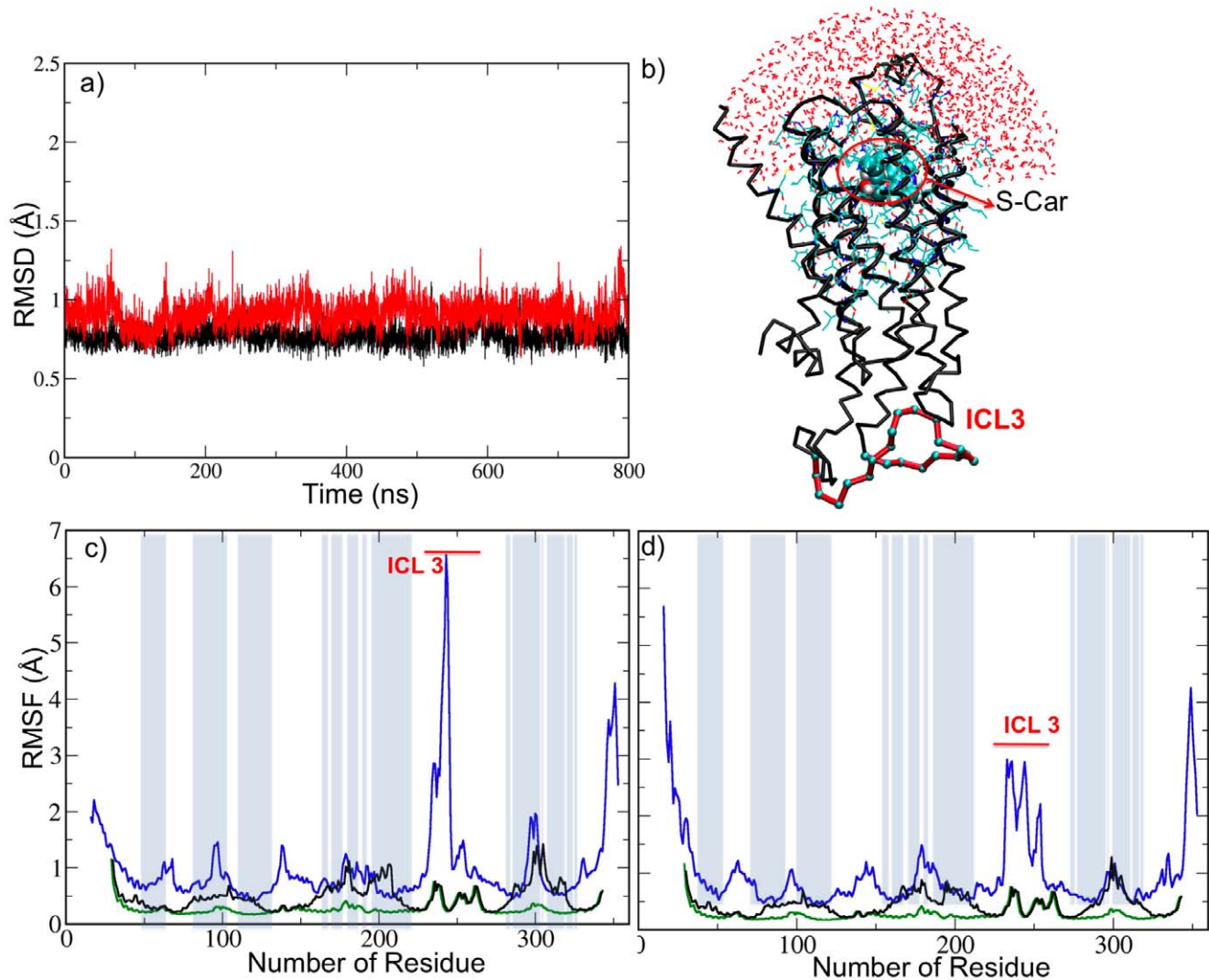
CG simulations were carried out for up to 1  $\mu$ s using a 2 fs time-step. Within this Go-model, each residue is treated as one bead at the position of the C $\alpha$  atom. Therefore the atom number of this system is 314, which is the number of residues present in X-ray structure.

## Results and Discussion

The accuracy of this method was established by performing MM/CG simulations of the human GPCR  $\beta_2$  adrenergic receptor (h $\beta_2$ -AR) in complex with its inverse agonist S-Carazolol (S-Car) and its agonist R-Isoprenaline (R-ISO), and by comparing these with a 800 ns all atom simulation of this system embedded in a lipid bilayer [19]. Both the all-atom MD and the MM/CG calculations are based on the h $\beta_2$ -AR/S-Car complex X-ray structure [25].

The MM/CG simulations were carried out for up to 800 ns. The MM region consisted of 476 and 486 atoms, while the overall system was made of 4597 and 4587 atoms for the h $\beta_2$ -AR.S-Car and h $\beta_2$ -AR.R-Iso complexes respectively. This allows us to run more than 80 ns/day on 16 CPUs (2.93 GHz Intel Xeon), a 15-fold speedup compared to MD simulations of the same system (Ursula Roethlisberger, private communication). With respect to this speed-up, we notice that a significant speed-up could also have been obtained using an all-atom representation for the whole protein while adding waters on both the intracellular and the extracellular region. However, the combined MM/CG representation allows a further reduction of the system size, since (i) waters are only needed in the extracellular site of the system; and (ii) the use of the CG representation for residues far away from the binding cavity reduces the system size.

Along the MM/CG trajectories, both complex structures remain stable, as observed in the root-mean-square deviation calculation (Figure 2a and Figure S4). The residue root-mean-square fluctuation calculated using the MM/CG approach follows

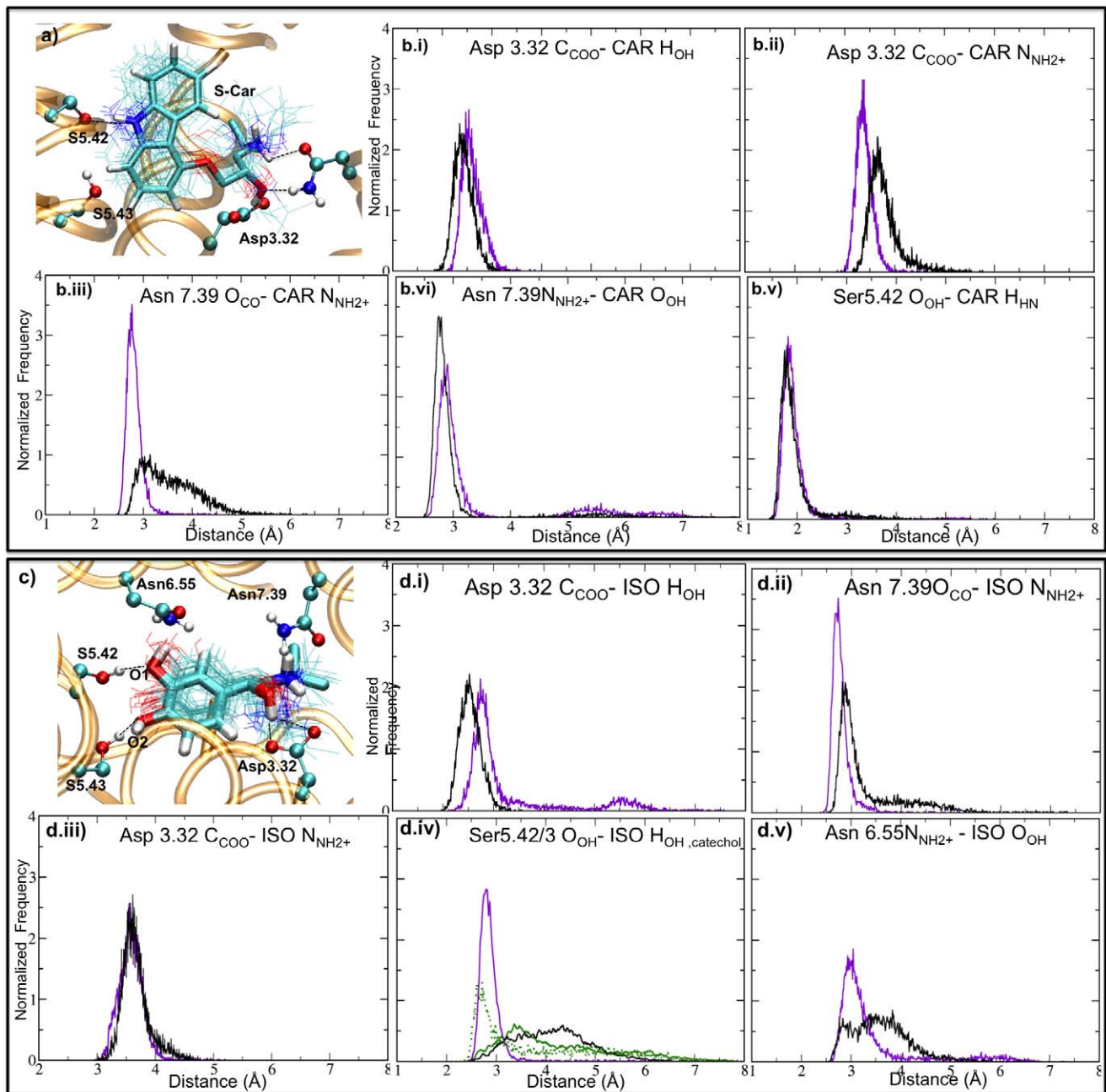


**Figure 2. Results of the simulations.** a) Root-mean-square-deviation of h $\beta$ 2-AR's backbone atoms in the MM/CG simulation of h $\beta$ 2-AR-S-Car (black lines) and h $\beta$ 2-AR-R-ISO (red lines), relative to the initial X-ray structure, plotted as a function of time. b) MM/CG representation of the h $\beta$ 2-AR-S-Car complex. The MM and I regions, together with the water molecules, are shown in a line representation, the ligand S-Car is shown as spheres, and the ICL3 is highlighted in red. c) and d) Root Mean-Square Fluctuations of h $\beta$ 2-AR's backbone atoms calculated based on MD simulations. Results for all-atom simulations [19], MM/CG simulations and CG simulations are shown in blue, black and green lines respectively. Results for h $\beta$ 2-AR-R-ISO and h $\beta$ 2-AR-S-Car complexes are shown in panels c) and d), respectively. Residues included in the MM and I regions (which feature all-atom representation) are highlighted by grey bars on the plots.  
doi:10.1371/journal.pone.0047332.g002

the trend obtained with all-atom MD (Figure 2c,d). The results in panels c) and d) show that globally the fluctuations of the MM and I regions are similar to the fluctuations observed in the all-atom MD simulations. In the CG regions, the fluctuations are much lower, possibly due to the higher rigidity of the CG Go-model force field in comparison to the all-atom force field [12]. This is observed in the full CG simulations (green line in Figure 2c,d). The main discrepancies between all atom MD and MM/CG simulations are observed in the intracellular loop 3 (ICL3, ranging from residues 233 to 253 approximately), which exhibits larger flexibility in the all-atom simulations. As shown in panel b), the ICL3, which is not present in the crystal structure [25], is located in the cytoplasmic region far away from the binding cavity. Thus, it was included in the CG region, and it is not expected to directly affect the properties of the binding site. Small differences in the fluctuations (smaller than 1Å) observed in the MM region with

respect to the all-atom simulation can be expected due to the different force fields applied in both simulations (Amber99 [32] and Gromos43a1 [15], respectively). The differences observed here are of the same order of magnitude as those observed in other comparisons between different force fields [33]. Indeed, these differences were also observed in two other independent test simulations of the h $\beta$ 2-AR/S-Car complex (Figure S5). The velocity autocorrelation function and the radial distribution function for the water molecules (Figure S2, Table S2) do not differ significantly from the data obtained for bulk water [34,35]. This suggests that the MM/CG scheme does not significantly constraint either the structure or the mobility of water molecules.

The structural determinants at the active site were well maintained (Figure 3). The interactions between the ligands and the protein matrix evolved in good agreement with the all-atom simulations. These included H-bond interactions between the



**Figure 3. MM/CG and MD simulations of  $h\beta 2$ -AR-S-Car and  $h\beta 2$ -AR-R-Iso complexes.** H-bond interactions between S-Car and  $h\beta 2$ -AR, reported in ref. [19]. Panels a) and c) display snapshots of the binding site, obtained from the MM/CG trajectory of the S-Car and R-Iso complexes respectively. Superimposed positions of the agonist and reverse agonist along the trajectories are shown in lines representation (snapshots were taken every 40 ns). The distribution of all H-bonds and salt involved in S-Car and R-Iso binding are shown in Panels b.i) to b.v) and d.i) to d.v). Results of the MM/CG and the all-atom MD simulations [19] are shown in black and violet lines respectively. In d.iv), black and violet lines correspond to the distance between the O1 (labeled in panel a) and the OH group of Ser5.42 in the MM/CG simulations, and the O2 and the OH group of Ser5.42 in the all-atom simulations, respectively. The distance between O1 and Ser5.43-OH and between O2 and Ser5.43-OH in the MM/CG simulation are shown in full and dotted green lines respectively.  
doi:10.1371/journal.pone.0047332.g003

ligands and residues Asn7.39, Asp3.32, Ser5.42, Ser5.43 and Asn6.55 (the last two residues, only in the case of the  $h\beta 2$ -AR-R-Iso complex). However, small differences have been observed in some specific interactions. The H-bonds between Asn7.39 carbonyl group ( $O_{CO}$ ) and the  $NH_2^+$  group of the agonist (R-Iso) or the reverse agonist (S-car) are only partially formed in the MM/CG simulation. (Figure 3 panel b.iii, d.ii). The H-bond

between  $NH_2^+$  group of Asn6.55 and R-Iso.OH is longer in the MM/CG simulations than in the all-atom simulations (Figure 3, panel d.v). Finally, both R-Iso.OH groups form H-bonds to Ser5.42 and Ser5.43 in the MM/CG simulation, while only the R-Iso.O(2)H group forms an H-bond to Ser5.42 in the all-atom simulation (Figure 3, panel d.iv) [19]. The fact that the simulations performed in this work show a high level of agreement with the all-

atom simulations allows us to suggest that the results do not critically depend on the choice of the force field.

To investigate the predictive power of the method at the structural level, we ran an additional simulation in which we located the ligand S-Car in a position different from the crystallographic pose (Figure S3). In this new position, none of the interactions with the residues found in the X-ray structure of h $\beta$ 2-AR/S-Car complex [25] are present. In these new simulations, the ligand migrates to the correct pose between 150 and 200 ns, forming the key interactions with Ser5.42, Ser5.43, Asp 3.32 and Asn7.39 (Figure S3). Hence, our method is not only able to conserve the ligand-receptor structure but also able to predict the correct pose of the ligand in the binding site.

All together, the MM/CG simulations reproduce the key structural features of the active site found in the MD simulations. The introduction of the potential wall to represent the membrane leads to a large reduction in the computational cost of the simulation, conserving the stability of the protein structure. Moreover, the ligand remains in a stable position inside the binding cavity throughout the long-scale simulation, conserving the key interactions with the protein matrix at the binding site. Due to its low cost and high accuracy, this method can be applied in this context to the study of a large number of GPCRs-ligand complexes.

## Conclusions

We have presented a hybrid MM/CG method to investigate h $\beta$ 2-AR, a receptor from the GPCR superfamily. The method allows a large speed up of the simulation while conserving all the key information related to the ligand-receptor interaction. This method can be extended to a large number of GPCR/ligand complexes and may be very useful in computer-aided drug design. Our code is implemented in GROMACS 4.5 [31] and is freely available upon request.

Combining MM/CG with model-built structures from homology modeling and/or molecular docking (such as in ref. [36]) may be a useful tool for structural predictions of GPCR/ligand complexes. The method allows the efficient and relatively cheap sampling of side chains orientations at the binding site while fully including hydration. This is particularly important for GPCRs, as water molecules can be found at the binding site of these receptors, waters which may be crucial to stabilizing the ligand [37,38].

## Supporting Information

**Figure S1** Wall potentials. (A) Schematic of the wall potentials  $V_i(d)$  ( $i = 1, 2$ ) plotted as a function of distance to the walls  $d$ . (B) The membrane wall  $\phi_5$  and the potential  $V_5(d)$  shifted by  $2r_p$ . (TIFF)

**Figure S2** Velocity autocorrelation (C(t)) function for the oxygen atoms of the water molecules in the MM/CG simulation of the hB2-AR/S-Car complex. The correlation of the velocities is lost after  $\sim 0.6$  ps, in agreement with the results previously obtained for a solution of SPC waters [35]. (TIFF)

## References

- Römpler H, Stäubert C, Thor D, Schulz A, Hofreiter M, et al. (2007) G protein-coupled time travel: evolutionary aspects of GPCR research. *Mol Interv* 7: 17–25.
- Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5: 993–996.
- Yarnitzky T, Levit A, Niv MY (2010) Homology modeling of G-protein-coupled receptors with X-ray structures on the rise. *Curr Opin Drug Discov Devel* 13: 317–325.
- Sgourakis NG, Garcia AE (2010) The membrane complex between transducin and dark-state rhodopsin exhibits large-amplitude interface dynamics on the

**Figure S3** MM/CG simulation of hB2-AR/S-Car complex. Here the S-Car ligand is originally located at a position different from the crystallographic pose. Panels a and b show snapshots taken at 0 ns and 180 ns of the simulation. Panel c shows the RMSD of the S-Car ligand with respect to the crystallographic position.

(TIFF)

**Figure S4** Root-mean-square-deviation per residue of h $\beta$ 2-AR's backbone atoms in the MM/CG simulation of a) h $\beta$ 2-AR.S-Car and b) h $\beta$ 2-AR.R-ISO relative to the initial X-ray structure. Overall, the protein's residues remain close to the crystal structure, with an RMSD lower than  $\sim 2$  Å. The regions with higher fluctuations with respect to the crystal structure consist of residues  $\sim 158$  to 200 (in helix IV), and  $\sim 300$ –305 (N-terminal extreme of helix VII), in agreement with the results presented in Figure 2c,d. These regions also show fluctuations in the all-atom simulations (as observed in Figures 2c,d), and do not include any of the residues interacting directly with the ligand in the binding cavity.

(TIFF)

**Figure S5** Root-mean-square fluctuation (RMSF) of the backbone atoms calculated for two independent 400 and 800 ns simulations of the hB2-AR/S-Car complex, shown with cyan and red lines respectively. For comparison purposes, the RMSF calculated for the all-atom simulation and for the MM/CG simulation described in the main text are shown with blue and black lines respectively. Grey bars indicate the MM and I regions. No large differences among the MM/CG simulations are observed.

(TIFF)

**Table S1** Available GPCR's structures in Protein Data Bank (PDB), adapted from <http://blanco.biomol.uci.edu/mpstruc> on 14 April 2012.

(DOC)

**Table S2** Maximum and minimum positions (in Å) of the Oxygen-Oxygen radial distribution functions for the oxygen atoms of the water molecules in the MM/CG simulation of the hB2-AR/S-Car complex. The values of the positions obtained for a solution of SPC waters are shown for comparison, which are taken from reference [34].

(DOC)

**Methods S1** Details on the MM/CG parameters (DOC).

(DOC)

## Acknowledgments

We thank Ursula Rothlisberger for kindly providing the trajectories of the all-atom MD simulations, and Justin Finnerty and Alexander Peyser for comments on the manuscript. Computer resources provided by the Jülich Supercomputing Center on JUROPA and by JARA-HPC on RWTH Compute Cluster (project ID jara0023) is greatly acknowledged.

## Author Contributions

Conceived and designed the experiments: AG PC ZH. Performed the experiments: ML CN LC. Analyzed the data: LC. Wrote the paper: ML CN LC ZH AG PC.

- sub-microsecond timescale: insights from all-atom MD simulations. *J Mol Biol* 398: 161–173.
5. Ayton GS, Noid WG, Voth GA (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. *Curr Opin Struct Biol* 17: 192–198.
  6. Rompler H, Staubert C, Thor D, Schulz A, Hofreiter M, et al. (2011) Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* 2: 17–25.
  7. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, et al. (2008) The MARTINI coarse-grained force field: Extension to proteins. *J Chem Theory Comput* 4: 819–834.
  8. Kalli AC, Campbell ID, Sansom MSP (2011) Multiscale simulations suggest a mechanism for integrin inside-out activation. *Proc Natl Acad Sci USA* 108: 11890–11895.
  9. Messer BM, Roca M, Chu ZT, Vicatos S, Kilshtain AV, et al. (2010) Multiscale simulations of protein landscapes: using coarse-grained models as reference potentials to full explicit models. *Proteins* 78: 1212–1227.
  10. Shi Q, Izvekov S, Voth GA (2006) Mixed atomistic and coarse-grained molecular dynamics: Simulation of a membrane-bound ion channel. *J Phys Chem B* 110: 15045–15048.
  11. Villa E, Balaef A, Schulten K (2005) Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc Natl Acad Sci USA* 102: 6783–6788.
  12. Neri M, Anselmi C, Cascella M, Maritan A, Carloni P (2005) Coarse-Grained Model of Proteins Incorporating Atomistic Detail of the Active Site. *Phys Rev Lett* 95: 218102.
  13. Neri M, Baaden M, Carnevale V, Anselmi C, Maritan A, et al. (2008) Microseconds dynamics simulations of the outer-membrane protease T. *Biophys J* 94: 71–78.
  14. Rzepiela AJ, Louhivuori M, Peter C, Marrink SJ (2011) Hybrid simulations: combining atomistic and coarse-grained force fields using virtual sites. *Phys Chem Chem Phys* 13: 10437–10448.
  15. Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, et al. (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* 26: 1719–1751.
  16. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25: 1656–1676.
  17. Go N, Abe H (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* 20: 991–1011.
  18. Deupi X, Standfuss J (2011) Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr Opin Struct Biol* 21: 541–551.
  19. Vanni S, Neri M, Tavernelli I, Rothlisberger U (2011) Predicting novel binding modes of agonists to  $\beta$  adrenergic receptors using all-atom molecular dynamics simulations. *PLoS Comput Biol* 7: e1001053.
  20. Taylor MRG (2007) Pharmacogenetics of the human beta-adrenergic receptors. *Pharmacogenomics J* 7: 29–37.
  21. Hoffmann C, Leitz MR, Oberdorf-Maass S, Lohse MJ, Klotz K-N (2004) Comparative pharmacology of human beta-adrenergic receptor subtypes—characterization of stably transfected receptors in CHO cells. *Naunyn Schmiedeberg Arch Pharmacol* 369: 151–159.
  22. Abe H, Go N (1981) Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. *Biopolymers* 20: 1013–1031. doi:10.1002/bip.1981.360200512.
  23. Nadler W, Brünger AT, Schulten K, Karplus M (1987) Molecular and stochastic dynamics of proteins. *Proc Natl Acad Sci USA* 84: 7933–7937.
  24. Osher S, Sethian J (1988) Fronts Propagating with Curvature-Dependent Speed - Algorithms Based on Hamilton-Jacobi Formulations. *J Comput Phys* 79: 12–49.
  25. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, et al. (2007) High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein Coupled Receptor. *Science* 318: 1258–1265.
  26. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779–815.
  27. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B, editor. *Intermolecular Forces*. D. Reidel Publishing Company, 331–342.
  28. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23: 327–341.
  29. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97: 10269–10280.
  30. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21: 1049–1074.
  31. Kutzner C, van der Spoel D, Fechner M, Lindahl E, Schmitt UW, et al. (2007) Speeding up parallel GROMACS on high-latency networks. *J Comput Chem* 28: 2075–2084.
  32. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, et al. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 117: 5179–5197.
  33. Rueda M, Ferrer-Costa C, Meyer T, Pérez A, Camps J, et al. (2007) A consensus view of protein dynamics. *Proc Natl Acad Sci USA* 104: 796–801.
  34. Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A* 105: 9954–9960.
  35. Paesani F, Zhang W, Case DA, Cheatham TE, Voth GA (2006) An accurate and simple quantum model for liquid water. *J Chem Phys* 125: 184507.
  36. Biarnés X, Marchiori A, Giorgetti A, Lanzara C, Gasparini P, et al. (2010) Insights into the binding of Phenyltiocarbamide (PTC) agonist to its target human TAS2R38 bitter receptor. *PLoS ONE* 5: e12394.
  37. Angel TE, Chance MR, Palczewski K (2009) Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc Natl Acad Sci USA* 106: 8555–8560.
  38. Nygaard R, Valentin-Hansen L, Mokrosinski J, Frimurer TM, Schwartz TW (2010) Conserved Water-mediated Hydrogen Bond Network between TM-I, -II, -VI, and -VII in 7TM Receptor Activation. *J Bio Chem* 285: 19625–19636.