



Published in final edited form as:

J Comput Chem. 2016 February 5; 37(4): 416–425. doi:10.1002/jcc.24231.

DIRECT-ID: An Automated Method to Identify and Quantify Conformational Variations - Application to β_2 -adrenergic GPCR

Sirish Kaushik Lakkaraju, Justin A. Lemkul, Jing Huang, and Alexander D. MacKerell Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland Baltimore, MD 21201

Abstract

The conformational dynamics of a macromolecule can be modulated by a number of factors, including changes in environment, ligand binding and interactions with other macromolecules, among others. We present a method that quantifies the differences in macromolecular conformational dynamics and automatically extracts the structural features responsible for these changes. Given a set of molecular dynamics (MD) simulations of a macromolecule, the norms of the differences in covariance matrices are calculated for each pair of trajectories. A matrix of these norms thus quantifies the differences in conformational dynamics across the set of simulations. For each pair of trajectories, covariance difference matrices are parsed to extract structural elements that undergo changes in conformational properties. As a demonstration of its applicability to biomacromolecular systems, the method, referred to as DIRECT-ID, was used to identify relevant ligand-modulated structural variations in the β_2 -adrenergic (β_2 AR) G-protein coupled receptor. Micro-second MD simulations of the β_2 AR in an explicit lipid bilayer were run in the apo state and complexed with the ligands: BI-167107 (agonist), epinephrine (agonist), salbutamol (long-acting partial agonist) or carazolol (inverse agonist). Each ligand modulated the conformational dynamics of β_2 AR differently and DIRECT-ID analysis of the inverse-agonist vs. agonist-modulated β_2 AR identified residues known through previous studies to selectively propagate deactivation/activation information, along with some previously unidentified ligand-specific microswitches across the GPCR. This study demonstrates the utility of DIRECT-ID to rapidly extract functionally relevant conformational dynamics information from extended MD simulations of large and complex macromolecular systems.

Keywords

conformational dynamics; covariance analysis; norm of matrix; ligand efficacy; microswitches; β_2 -adrenergic G-protein coupled receptor

Introduction

Knowledge about macromolecular conformational dynamics is essential for understanding biological function.¹ Molecular dynamics (MD) simulations are extremely powerful towards

*Corresponding author, alex@outerbanks.umaryland.edu, Address: 20 Penn St., Room 633, Baltimore, MD 21201, Phone: (410) 706-7442, Fax: (410) 706-5017 .

this understanding in that they provide high-resolution details of spatial arrangement of particles in a system over time. Atomistic MD simulations spanning a wide range of time-scales have been employed extensively in the past to understand different aspects of dynamics and functions of biomacromolecules of different sizes.²⁻⁵ However, the overwhelming amount of information in MD trajectories detailing the time-evolution of the system is often under-utilized.⁶ Thus, it is beneficial to refine analysis methods to optimize the extraction of functionally relevant conformational information.

Covariance matrices quantify coupling between atomic displacements and provide a detailed means of extracting comprehensive information from MD trajectories. Information from the covariance matrices has been useful for quantifying configurational entropy,^{7,8} identifying quasi-harmonic modes,⁹ and extracting functionally relevant global motions.¹⁰ In Cartesian space, for an N -atom system, a covariance matrix contains $3N \times 3N$ elements, making extraction of relevant information challenging for large systems, given the high dimensionality. In Principal Component Analysis (PCA),¹¹ the covariance matrices are diagonalized to yield collective variables (eigenvectors) sorted according to their contribution to the total mean-squared fluctuation.⁶ In other words, the MD trajectory is projected onto principal components (collective variables), and it has been found that these collective coordinates constitute a low-dimensional subspace that represents a significant part of the macromolecular motion.¹² Collective coordinates thereby drastically reduce the number of descriptors that are required to characterize the dynamics of the macromolecule.¹³

In certain cases, large-amplitude functionally relevant deformations are harmonic, aligning with the macromolecular intrinsic normal modes.^{14,15} Consequently, several studies have attempted to project macromolecular dynamics using elastic network models (ENM),^{16,17} where the macromolecule is defined as a set of nodes connected by harmonic springs.¹⁸ Normal modes/collective coordinates are then derived from this simplistic harmonic approximation of the potential energy function near equilibrium. Covariance-based methods delineating modes from ENM have been used to identify domains within a macromolecule that cluster structural features with concerted structural fluctuations.¹⁹⁻²¹ While this is functionally relevant information,^{22,23} ENM-based methods suffer from two major limitations. First, by either treating the macromolecule in isolation, or using implicit solvent schemes to screen inter-atomic interactions, the methods likely fail to account for configurational entropy of the surrounding environment. Second, given the ruggedness of the potential energy surface and anharmonicity of the energy barriers for conformational transitions, harmonic approximations tend to lose critical information relevant for characterizing macromolecular conformational behavior.^{14,24}

A number of factors modulate the conformational dynamics of a macromolecule. These include changes in ion^{25,26} or osmolyte^{27,28} concentration, ligand binding or dissociation,²⁹⁻³¹ and interactions with other macromolecules,³²⁻³⁴ to list a few. While some of these factors modulate large domain motions,^{35,36} others are more subtle.^{37,38} Often there is a large overlap in the low-frequency global motions across these different conformational equilibria, with only small changes occurring in the high frequency motions. All of these

effects are likely to be missed when characterizing the dynamics through ENM or normal mode analysis.

In the present work, we describe a method called DIRECT-ID (Dimensionality Reduction through Covariance matrix Transformation for Identification of Differences in dynamics) that quantifies changes in the macromolecular conformational dynamics using the norm of the differences in the covariance matrices calculated from multiple MD trajectories. The MD simulations can be independent runs under different conditions to gain insight into the modulation of macromolecular conformational dynamics. For each pair of trajectories, structural features with the largest changes in atomic motions are identified. The covariance matrices are constructed from non-hydrogen atoms of the macromolecule, thus the method is sensitive to both global motion changes as well as smaller higher frequency mode changes. Given that the bonds involving hydrogen atoms are generally constrained in most biomolecular MD simulations, their dynamics are dominated by the associated non-hydrogen atoms and thus not included in the current implementation of DIRECT-ID.

The ability of DIRECT-ID to identify functionally relevant features is evaluated by probing the modulation of conformational dynamics of the β_2 -adrenergic G-protein coupled receptor (β_2 AR) by ligands of different functions. Studies have shown that instead of a previously hypothesized two state active-inactive model, GPCRs have multiple, distinct active-inactive conformations.³⁹ Ligands of different efficacies modulate the conformations of GPCRs differently,⁴⁰ ultimately affecting the downstream functional signaling profiles. Consequently, a number of computational^{41,42} and experimental³¹ studies have probed the effect of ligand binding on β_2 AR. Thus, the β_2 AR is a good system to validate the ability of DIRECT-ID to quantify changes in the conformational dynamics and identify structural features that are sensitive to these changes.

Method

1) Multiple MD simulations of the macromolecule of interest are initially performed. These simulations may vary in any manner dictated by the user. Each of these simulations is recommended to span similar time-scales and should be tested for convergence with any of the conventionally used metrics (e.g., RMSD of the backbone atoms) or using our convergence analysis (see below). From the trajectories, the Cartesian coordinates of the atoms of interest are used for the calculation of the covariance matrices and the analysis as follows.

2) In the Cartesian space, the covariance between two atoms, along an axis x , is calculated as $\langle r1_x \cdot r2_x \rangle$, where $r1_x = r1_x - r1_{0x}$, and $r2_x = r2_x - r2_{0x}$, $r1_x$ and $r1_{0x}$, $r2_x$ and $r2_{0x}$ are the positions at a time t and in a reference conformation along an axis x for atoms 1 and 2, respectively. Angle brackets denote ensemble average. For a system containing N atoms being analyzed, the covariance matrix is symmetric with $3N \times 3N$ elements,

$$\mathbf{C} = \begin{bmatrix} \langle \Delta r_{1x} \cdot \Delta r_{1x} \rangle & \cdots & \cdots & \langle \Delta r_{1x} \cdot \Delta r_{Nz} \rangle \\ \vdots & \langle \Delta r_{1y} \cdot \Delta r_{1y} \rangle & & \vdots \\ \vdots & & \ddots & \vdots \\ \langle \Delta r_{Nz} \cdot \Delta r_{1x} \rangle & & & \langle \Delta r_{Nz} \cdot \Delta r_{Nz} \rangle \end{bmatrix}$$

3) Given a set of MD simulations, i, j, \dots, M of the N -atom system, the covariance matrices $\mathbf{C}_i, \mathbf{C}_j, \dots, \mathbf{C}_M$ are calculated with respect to a single reference configuration, \mathbf{r}_0 .

4) Differences between two covariance matrices \mathbf{C}_i and \mathbf{C}_j are calculated as

$$\mathbf{D}_{i-j} = \begin{bmatrix} \left| \langle \Delta r_{1x} \cdot \Delta r_{1x} \rangle_i - \langle \Delta r_{1x} \cdot \Delta r_{1x} \rangle_j \right| & \cdots & \cdots & \left| \langle \Delta r_{1x} \cdot \Delta r_{Nz} \rangle_i - \langle \Delta r_{1x} \cdot \Delta r_{Nz} \rangle_j \right| \\ \vdots & \left| \langle \Delta r_{1y} \cdot \Delta r_{1y} \rangle_i - \langle \Delta r_{1y} \cdot \Delta r_{1y} \rangle_j \right| & & \vdots \\ \vdots & & \ddots & \vdots \\ \left| \langle \Delta r_{Nz} \cdot \Delta r_{1x} \rangle_i - \langle \Delta r_{Nz} \cdot \Delta r_{1x} \rangle_j \right| & & & \left| \langle \Delta r_{Nz} \cdot \Delta r_{Nz} \rangle_i - \langle \Delta r_{Nz} \cdot \Delta r_{Nz} \rangle_j \right| \end{bmatrix}$$

It is important that the same reference configuration \mathbf{r}_0 be used for the calculation of the covariance matrices from the different simulations, instead of the mean structures from within the simulations. This is because, when the same \mathbf{r}_0 is used in calculation of \mathbf{C}_i , and \mathbf{C}_j across two trajectories i and j , then elements of the \mathbf{D}_{i-j} matrix,

$D_{pq} = \langle r_p^i r_q^i \rangle - \langle r_p^j r_q^j \rangle - r_p^0 (\langle r_q^i \rangle - \langle r_q^j \rangle) - r_q^0 (\langle r_p^i \rangle - \langle r_p^j \rangle)$, (where p and q refer to a row and column, respectively) characterize the differences in the atomic displacements between the two trajectories. On the other hand, when the mean structures from the simulations are used in \mathbf{C}_i , then $D_{pq} = \langle r_p^i r_q^i \rangle - \langle r_p^i \rangle \langle r_q^i \rangle - \langle r_p^j r_q^j \rangle - \langle r_p^j \rangle \langle r_q^j \rangle$, characterize the differences in the variance of the motion.

5) Differences in conformational dynamics across two trajectories are quantified using the norm of the difference matrix. Norm, $\|\mathbf{D}_{i-j}\|_1$, is the maximum absolute column sum of the difference matrix and can be considered as the magnitude of the matrix. A higher norm signifies a greater difference between the two corresponding covariance matrices. Given the symmetric nature of the covariance and difference matrices, calculations are invariant between $\|\mathbf{D}_{i-j}\|_1$ and $\|\mathbf{D}_{i-j}\|_\infty$, the maximum absolute row sum. Thus, for M trajectories, we have a matrix of norms,

$$\mathbf{N}_{M \times M} = \begin{bmatrix} \|\mathbf{D}_{i-i}\|_1 & \cdots & \cdots & \|\mathbf{D}_{i-M}\|_1 \\ \vdots & \|\mathbf{D}_{j-j}\|_1 & & \vdots \\ \vdots & & \ddots & \vdots \\ \|\mathbf{D}_{M-i}\|_1 & & & \|\mathbf{D}_{M-M}\|_1 \end{bmatrix}$$

that is a triangular matrix with diagonal elements equal to 0.

6) Thus, differences across M covariance matrices of $3N \times 3N$ elements are transformed into a single norm matrix containing $M \times M$ elements. $\mathbf{N}_{M \times M}$ can be parsed to identify trajectories with the biggest differences in their conformational dynamics. Difference matrices, \mathbf{D}_{i-j} , between each pair of trajectories, i and j , can be parsed to identify elements with the highest magnitude. The elements correspond to pairs of non-hydrogen atoms in the macromolecule with largest changes in the correlated motions between the two trajectories. These atom pairs are then used to establish a list of residues/nucleotides (depending on the nature of the macromolecule) associated with the corresponding non-hydrogen atoms; an example is shown in Table 1. Thus, the final output is a table that lists macromolecular residues with the most prominent changes in their dynamics between the two trajectories.

DIRECT-ID is written as a C++ program that can handle covariance matrices outputted from CHARMM⁴³ or the GROMACS⁴⁴ utility “*gmx covar*.” The code extracts structural features using these matrices. Note that in general any molecular simulation software can be used to generate the trajectories and calculate the covariance matrices, as the C++ code only requires as ASCII covariance matrix of the form shown above in step 2.

Simulation Details

One micro-second MD simulations of the β_2 AR were run in both the apo form and in complex with ligands: BI-167107 (BIA),³⁵ epinephrine (EPNE),⁴⁵ salbutamol (SALB),⁴⁶ and carazolol (CARA).⁴⁷ The active conformation of β_2 AR from PDB 3P0G³⁵ was used to initiate the MD simulations. The region of the structure corresponding to T4 lysozyme, which was inserted between TM5 and TM6 to facilitate crystallization, was deleted and the third intracellular loop (ICL3) was built in its place using MODELLER⁴⁸. In total, 100 models were generated and ranked using the Discrete Optimized Protein Energy (DOPE) method⁴⁹ and the highest ranking model was used as a starting structure. Binding poses of ligands EPNE and CARA were obtained from PDB 4LDO and 2RH1, respectively. SALB was docked using Autodock-Vina,⁵⁰ and the top scoring binding pose used in the simulation had an orientation similar to that of EPNE. Disulphide bridges between residues 106-191 and 184-190 were maintained. All simulations were initiated from the same starting conformation which was also used as the reference in the DIRECT-ID analysis.

Empirical force field parameters for β_2 AR and lipids were taken from CHARMM36,⁵¹⁻⁵³ ligand topologies and parameters were derived from CGenFF⁵⁴ and water was treated using the TIP3P model.^{55,56} β_2 AR was inserted into a rectangular palmitoyl oleoyl phosphatidyl choline (POPC) membrane containing ~252 lipids, including 10% cholesterol using CHARMM-GUI;⁵⁷ the remainder of the unit cell volume was filled with ~0.15 M NaCl aqueous solution including sufficient counterions to neutralize the system. All systems were energy minimized in two steps, first by 1500 steps of steepest descent (SD)⁵⁸ and the next 1500 steps with the adopted basis Newton-Raphson (ABNR) algorithm.⁵⁹ During minimization, the non-hydrogen atoms of the protein backbone and side chains were harmonically restrained using force constants of 10 kcal/mol/Å² and 5 kcal/mol/Å², respectively. Influx of water molecules into the hydrophobic core of the TM region was prevented using a harmonic restraining potential of 2.5 kcal/mol/Å² along the x-y plane at positions of ± 11 Å along the z-axis from the center of the receptor. The same force was also

used to keep the heads and tails of the lipids in place, and configurations of the lipids were maintained with harmonic dihedral restraints with a force constant of 250 kcal/mol/rad². The same restraints were used during the 375 ps of equilibration with a 1-fs time step, and the restraint forces were gradually reduced as described in our previous work.⁶⁰ The first 50 ps of equilibration was performed using Langevin dynamics with a friction coefficient of 3 ps⁻¹ on all non-hydrogen atoms. The remaining 325 ps of the equilibration used constant pressure-temperature (NPT) dynamics using the Langevin Piston integrator.⁶¹ Covalent bonds involving hydrogens were fixed at the equilibrium length by the SHAKE algorithm.⁶² Lennard-Jones (LJ) interactions were switched to zero from 10 to 12 Å and the non-bonded pair list was generated out to 14 Å and updated heuristically.⁶³ Electrostatic interactions were calculated by the particle-mesh Ewald⁶⁴ summation method with a real space cutoff of 12 Å. These minimization and equilibration simulations were performed using CHARMM.⁵⁹

Following equilibration, production runs were performed using GROMACS⁴⁴ and with the leapfrog integrator (GROMACS integrator “md”) with a time step of 2 fs. The Nosé–Hoover method^{65,66} was used to maintain the temperature at 300 K with the protein and the remainder of the system separately coupled to heat baths. Pressure was maintained at 1 bar using the Parrinello–Rahman⁶⁷ barostat. The time constant used for temperature and pressure coupling were both set to 1 ps. The LINCS^{68,69} algorithm was used to constrain all covalent bonds involving hydrogen atoms and water geometries were kept rigid with SETTLE.⁷⁰ LJ interactions were switched off smoothly in the range of 10–12 Å, and the particle mesh Ewald (PME) method^{64,71} was used to treat long-range electrostatics with cubic B-spline interpolation and a maximum grid spacing of 1.2 Å. Long-range dispersion correction was applied to the energy and pressure terms. No restraints were applied during the production.

Results

Across the five 1- μ s MD simulations, the overall structure of the β_2 AR was stable, with the root mean square deviation (RMSD) of the backbone atoms across the helices H1-H8 (excluding the loops), plateauing at around 0.4 μ s (SI, Fig. S1). Ligands were also stable in the β_2 AR binding pocket, maintaining interactions with the residues in helices H3, H5, H6 and H7 (SI, Fig. S2). EPNE was slightly more flexible, with a deviation from the crystallographic conformation modulated by a break in its hydrogen bonding with N312^{7.38} (superscript indicates Ballesteros-Weinstein numbering), while preserving interactions with the rest of the binding pocket (SI, Fig. S2C). Despite the similarity in the backbone RMSD, important differences were noted in the conformational dynamics across the five simulations, as detailed below.

Ligand Modulated G-protein Binding Cavity in β_2 AR

Structural studies have identified two distinct conformations of β_2 AR, in which the opening and closing of this cavity are correlated with the activation and inactivation of the GPCR, respectively.^{35,72} Helix α_5 in G_{qs} of the G-protein heterotrimer enters into this cavity which is located at the intracellular side of the GPCR, and interactions between α_5 and the GPCR

are known to initiate the subsequent intracellular signaling downstream.^{73,74} To probe the heterogeneity in the conformational dynamics of β_2 AR in the apo state and with the different ligands, this G-protein binding cavity was traced using the distances between the center of mass of intracellular side of helices H2-H6 and H3-H7. Residues considered for the center of mass calculation were: V67-Y70(H2), S120-V129(H3), C265-H269(H6) and G320-S329(H7). Using these two distances as reaction coordinates, approximate free energy surfaces were calculated as $F(n) = -k_B T \ln(N_n)$, where k_B is the Boltzmann constant (kcal/mol/K), T is the temperature (300 K), and N_n is the number of times the grid point is visited during the length of a simulation. These reaction coordinates were based on previous conformational dynamics studies that probed GPCR-G-protein coupling specificity⁷⁵ and passage of water through a GPCR channel.⁷⁶

As shown in Fig. 1, conformational dynamics of β_2 AR are clearly different between its apo state and when in complex with different ligands. Apo and CARA-bound β_2 AR transitioned to distinct “inactive-like” conformations, in which the cavity that accommodates $\alpha 5$ of $G_{\alpha s}$ is not formed (Fig. 1B and C). Agonists, on the other hand, retained the G-protein cavity, but the size of this cavity was smaller than that found in the active crystal structures (Fig. 1D-F), in agreement with recent studies.^{75,76} Conformational sampling was also distinct between the agonists, such that BIA-bound β_2 AR more frequently sampled “active-like” conformations with bigger cavities, than when bound to EPNE or SALB. SALB-bound β_2 AR also visited more inactive-like conformations, indicating a need for a second ligand or a G-protein to stabilize the active conformation in the presence of this partial agonist. Consistent with previous studies,⁴² the opening of the cavity is found to be correlated with a break in the bond between R131 of H3 and E268 of H7 SI, Fig. S3). This ionic-bond formed most predominantly in the APO state of β_2 AR and not observed in the agonist-bound simulations. Notably, although R131-E268 can characterize the transitions between the end-states of β_2 AR, at least two reaction coordinates are needed to delineate heterogeneity in ligand modulated changes to conformational dynamics of β_2 AR. For instance, cavity opening can also be traced using distances between H3-H6 and H3-H7 (SI, Fig. S4). These surfaces also depict the conformational heterogeneity similar to that shown in Fig. 1.

DIRECT-ID Analysis

It is important to note that the choice of the reaction coordinates for the surfaces in Fig. 1 was based on knowledge derived from previous structural studies. These surfaces describe the more global domain-like motions in the receptor without any information about conformational changes of individual residues. Recent studies have shown that conformational changes in a subset of residues, often called “microswitches”, manifest into macroscopic conformational changes indicative of activation/inactivation of the receptor.^{77,78} To characterize the changes in the conformational equilibria without prior knowledge of reaction coordinates, and to identify structural features such as the microswitches that are responsive to these changes, DIRECT-ID analysis was performed using the trajectories from the five MD simulations. For each simulation, covariance of the atomic displacements was calculated over the 1- μ s trajectory for all the non-hydrogen atoms in the seven transmembrane helices (excluding loops), yielding five covariance matrices that each contained 3879×3879 elements. Differences between these covariance matrices

contain all the variations in conformational dynamics, but given the size of these systems, representation of these matrices becomes intractable (see Fig. 2A). Thus the differences are quantified using $\|\mathbf{D}_{i-j}\|_1$ as shown in Fig. 2B. The largest differences all involve the agonist BIA with the all other states, with the smallest difference occurring between the apo state and the inverse-agonist CARA, and the partial agonist SALB. The values of $\|\mathbf{D}_{i-j}\|_1$ are included in Table 1.

Ligand Efficacy using DIRECT-ID

As agonists maximally activate the receptor, and activation is coupled with conformational transition from the apo state, efficacy can be estimated from $\|\mathbf{D}_{\text{agonist-apo}}\|_1$. As shown in Fig. 2B, $\|\mathbf{D}_{\text{BIA-APO}}\|_1$ (403.1) > $\|\mathbf{D}_{\text{EPNE-APO}}\|_1$ (233.6) > $\|\mathbf{D}_{\text{SALB-APO}}\|_1$ (140.2). This outcome is consistent with previous observations that BIA has a greater efficacy than the natural ligand, EPNE, and the partial agonist induced submaximal activity even at saturating concentrations.³⁵ Thus, DIRECT-ID is able to quantify the effect of changes to the global motion of the protein, which is critical to the understanding of effects of different ligands.

Atomic Details of Conformational Variations

Once the systems with the largest differences are identified using the $\mathbf{N}_{M \times M}$ matrix, the corresponding \mathbf{D}_{i-j} matrices may be parsed to identify structural features (residues) with the most prominent changes between the two trajectories. Table 1 lists the residues thus obtained across all the pairs of the trajectories. Notably, residues that respond to changes in ligand properties are not limited to the binding pocket or helices H5 and H6, rather they are distributed all across the receptor as described in the following sections.

Agonists Mediated Differences

Distinct differences were found between conformational dynamics of $\beta_2\text{AR}$ when in complex with agonists versus the inverse-agonist, CARA. This is particularly significant when comparing dynamics between BIA- and CARA-bound $\beta_2\text{AR}$ (Fig. 2A,B). Structural features with the biggest changes in their motion between these two simulations span helices H3, H4, H6 and H7 (Fig. 2C, Table 1). Several of these residues (E122^{3,41}, T164^{4,56}, Y326^{7,53}) have been identified through previous network information analysis methods^{78,85} and site-directed mutational studies (see Table 1) to behave as microswitches and participate in communication pathways that relay activation/inactivation information from the ligand binding pocket to the intracellular G-protein binding regions. In addition, DIRECT-ID identified residues that selectively responded to changes in conformational equilibria of the receptor mediated by only some of the ligands and not others. For instance, W158^{4,50} underwent a conformational transition that involved a χ_2 -rotation, pushing it out of the largely hydrophobic cavity between the helices H2 and H3, only when $\beta_2\text{AR}$ was bound to BIA (Fig. 3A). Similarly, displacement of Q197^{5,37} that broke its interaction with H6 was also observed only in the BIA-bound $\beta_2\text{AR}$ (Fig. 3B). Y326^{7,53}, is a known microswitch, whose conformational transition allows passage of water through another family A GPCR, the adenosine receptor, was modulated only by its agonist.⁷⁶ Consistent with this mechanism, only in the BIA-bound $\beta_2\text{AR}$ did Y326 undergo a χ_1 -rotation that disrupted its hydrogen-bond network with H1, H2, and H6 and moved into an alternate rotameric

conformation that involved an aromatic stacking interaction with the conserved F332^{8.50} in TM8 (Fig. 3D).

In contrast, the presence of EPNE in the β_2 AR binding pocket modulated conformational transitions in N318^{7.45} that were not seen in BIA-bound β_2 AR (Fig. 3C). Specifically, interactions between N318 and helices H2, H3 and H6, and the N318^{7.45}-N322^{7.49} (another known microswitch) hydrogen bond were broken. Although occurring at different locations along the receptor, each of these transitions allow for greater flexibility within the transmembrane helices, resulting in a larger opening of G-protein binding cavities (Fig. 1D-F) that correlate with the activation of the receptor.

Partial-Agonist Mediated Differences

SALB is a long-acting partial agonist of β_2 AR. Functional studies have determined that SALB-mediated signaling is different from other agonists such as EPNE.⁸⁶ These changes were attributed to conformational transitions that significantly reduced GRK-modulated phosphorylation at the ICL3 and C-terminal tail of β_2 AR.^{87,88} Both ICL3 and the C-terminal tail were missing in the crystal structures and not included in the current analysis. Nevertheless, DIRECT-ID identified residues along the receptor with changes in their conformational motions selectively mediated by SALB-bound β_2 AR (Table 1). For example, F166^{4.58} (Fig. 4A) and Q197^{5.37} (Fig. 3B) had transitions similar to those seen in β_2 AR bound to agonists but different from those in the apo state or the CARA-bound β_2 AR. Others, such as the F217^{5.56} (Fig. 4B) underwent conformational transitions that were different from the agonists. χ_2 -rotation of F217 in SALB-bound β_2 AR disrupted its interaction with Y132^{3.51} moving into a hydrophobic cluster with L212^{5.51}, V213^{5.52} and V216^{5.55}. This transition likely creates a strain in H5 that could propagate downstream and modulate the phosphorylation of ICL3, although additional studies are needed to confirm this hypothesis. Notably, DIRECT-ID correctly identified these changes in side-chain conformations in SALB-bound β_2 AR, despite similarity in the more “global” conformational motions of H6 and H7 to inactive-like conformations seen in the apo state and CARA-bound β_2 AR (Fig. 1B,C vs. F).

DIRECT-ID analysis with the different ligands revealed the existence of multiple allosteric communication pathways/microswitches that selectively propagate information depending upon the state of the receptor. Some residues such as the R151^{4.43}, I153^{4.45}, and T164^{4.56} were consistently found to have differences in their conformational motion when comparing the apo state versus agonists bound β_2 AR. T164 is known to be polymorphic, with mutational changes modulating the receptor functional profile differently. T164V reduces basal activity and agonist-induced activation, while T164I was found to alter Salmeterol's (agonist) exosite binding and conventional agonist-driven coupling between GPCR and G_{α_s} of G-protein.⁸¹ Thus, changing the nature of the residue at a conserved position could modulate the communication pathway adopted and the consequent receptor functioning.

As a more thorough understanding of the effect of these residues on the signaling pathways emerges through future site-directed mutational studies, they (along with the others listed in Table 1) could be targeted using new allosteric modulators to bias the signaling profile of the receptor.

Convergence Assessment using DIRECT-ID

Resetting a fully activated β_2 AR into the inactivated state requires at least 100 ns,^{41,76} leading to recent MD simulations studies on the order of several microseconds in an effort to observe fully reversible transitions.^{76,89} To determine the time scales required to quantify the differences in conformational dynamics using DIRECT-ID, $\|\mathbf{D}_{i-j}\|_1$ was calculated over 0.1- μ s increments through the 1- μ s trajectories. As shown in Fig. 5A, differences in $\|\mathbf{D}_{i-j}\|_1$ converged at around 0.7 μ s for all the trajectories. Differences in local conformational transitions for residues identified using DIRECT-ID, such as the W158^{4,50} also converged at these time-scales (Fig. 5B). Thus, while longer MD simulations may be useful for fully characterizing reversible transitions across different states of the receptor, DIRECT-ID analysis demonstrates that the present simulations are sufficient for quantifying differences in conformational dynamics.

Discussion

Biased agonism is a novel and emerging concept, particularly in GPCR research.⁹⁰ Recent drug-design efforts have focused on developing ligands that stabilize a particular active conformation of the receptor and subsequently bias a beneficial signaling profile downstream. As the mechanistic link between macromolecular conformations and downstream signaling emerges, allosteric modulators could be developed to target residues identified using DIRECT-ID that selectively modulate a beneficial signaling profile. For instance, by targeting residues identified using DIRECT-ID to selectively respond to SALB, ligands could potentially reduce GRK-modulated phosphorylation of the ICL3 and the C-terminal tail, and thereby reduce β_2 AR desensitization, a known problem with the current drugs targeting β_2 AR.⁹¹ Additionally, DIRECT-ID will be useful towards both quantification of efficacies of the new ligands and identification of unique structural features that are responsive to the changes in conformational dynamics associated with these ligands.

The present results with the β_2 AR demonstrate the utility of DIRECT-ID in linking the heterogeneity of free energy surfaces delineating “global” changes (such the G-protein binding cavity in Fig. 1) with the more specific changes in the “local” residues associated with the conformational transitions (Fig. 3,4) without any loss of information. Specifically, this procedure allows for rapid identification of residues that contribute to the selective conformational changes of the receptor. For the 1- μ s trajectories, generation of covariance matrices using the *gmx covar* utility in GROMACS took ~6 min on a 12 core Xeon X5675 3.07 GHz CPU workstation. Extraction of structural features from a pair of these covariance matrices each containing 3879×3879 elements took ~1 min using DIRECT-ID. As the method operates with covariance matrices from MD simulations, anharmonicity of the conformational differences are implicitly taken into account. Dynamics across simulations can be restricted to a single global minimum or can span multiple macrostates. As the covariance matrices are defined based on the non-hydrogen atoms, even small conformational transitions such as the side-chain displacements will be identified. For instance, akin to the SALB-bound vs apo state β_2 AR, conformational landscapes between CARA-bound β_2 AR and the apo state are also similar (Fig. 1B,C) with the β_2 AR visiting inactive-like conformations through both the simulations. Consequently, $\|\mathbf{D}_{\text{CARA-APO}}\|_1$

(194.1) is smaller than $\|\mathbf{D}_{\text{agonists-APO}}\|_1$, thus quantifying the differences between the CARA-bound and the apo state $\beta_2\text{AR}$. Despite these similarities in conformational landscapes, DIRECT-ID was still able to extract structural features with more subtle changes in conformational transitions (Table 1), demonstrating its power in connecting microscopic and macroscopic dynamics that might otherwise go unnoticed.

A caveat of the DIRECT-ID method is that, if there is a structural shift, such as a change in the tertiary structure, without a change in the local conformational motions of elements of the secondary structure, then these elements may rank low in \mathbf{D}_{i-j} . However, it is important to note that such large-scale structural changes are unlikely to be independent of local coupled changes such as breakage of an ionic or hydrogen bond, disruption of hydrophobic packing, etc. Often the structural shift allows for greater conformational variability of the elements in the secondary structure. Each of these events, on the other hand, will rank high in \mathbf{D}_{i-j} , and hence will be extracted. Thus, DIRECT-ID will invariably identify a subset of structural features with significant changes in their conformational variability, and as demonstrated in the case of $\beta_2\text{AR}$, these structural features are likely to have critical roles in the modulating the function of the macromolecule.

Importantly, no prior knowledge of the reaction coordinates of the macromolecule is needed to study the modulation of its dynamics with DIRECT-ID. As the changes in the conformational motions across the non-hydrogen atoms of a macromolecule are reported, the method identifies structural features best suited for projecting the conformational dynamics of the macromolecule. It is important to note that the utility of DIRECT-ID is not restricted to GPCRs or even ligand-modulated changes in dynamics. The method is readily applicable to studying the modulation of conformational dynamics by a variety of factors and, as the results with the GPCR highlight, even very large systems can be analyzed. In its current implementation, from any set of M covariance matrices and a reference coordinate file in PDB format, the code outputs an $\mathbf{N}_{M \times M}$ matrix with $\|\mathbf{D}_{i-j}\|_1$ as its elements as shown in Fig. 2B, along with the list of residues as shown in Table 1. Given the overarching importance of characterizing the modulation of conformational properties of a macromolecule and the consequent implications on its functional profile, a wide applicability of this method is anticipated. Scripts and the code are made available at <http://mackerell.umaryland.edu>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grants GM051501, GM070855, and GM072558. J.A.L. is supported by NIH fellowship F32GM109632. The authors acknowledge computer time and resources from the Computer Aided Drug Design (CADD) Center at the University of Maryland, Baltimore, and the XSEDE supercomputing grid.

References

1. Karplus M. *Angew. Chem. Int. Ed.* 2014; 53:9992–10005.
2. McCammon JA, et al. *Nature.* 1977; 267:585–590. [PubMed: 301613]

3. Perilla JR, et al. *Curr. Opin. Struct. Biol.* 2015; 31:64–74. [PubMed: 25845770]
4. Shaw DE, et al. *Science*. 2010; 330:341–346. [PubMed: 20947758]
5. Lakkaraju SK, et al. *J. Chem. Info. Model.* 2013; 53:1337–1349.
6. Berendsen HJ, Hayward S. *Curr. Opin. Struct. Biol.* 2000; 10:165–169. [PubMed: 10753809]
7. Karplus M, Kushick JN. *Macromolecules*. 1981; 14:325–332.
8. Frappier V, Najmanovich R. *Protein Sci.* 2015; 24:474–483. [PubMed: 25367089]
9. Levy RM, et al. *Macromolecules*. 1984; 17:1370–1374.
10. Balsera MA, et al. *J. Phys. Chem.* 1996; 100:2567–2572.
11. Jolliffe, I. *Principal component analysis*. Wiley Online Library; 2002.
12. Amadei A, et al. *Proteins Structure Function and Bioinformatics*. 1994:412–425.
13. Kitao A, Go N. *Curr. Opin. Struct. Biol.* 1999; 9:164–169. [PubMed: 10322205]
14. Ma J. *Structure*. 2005; 13:373–380. [PubMed: 15766538]
15. Lakkaraju SK, Hwang W. *Phys. Rev. Lett.* 2009; 102:118102. [PubMed: 19392240]
16. Keskin O, et al. *Biochemistry*. 2002; 41:491–501. [PubMed: 11781087]
17. Bahar I, et al. *Curr. Opin. Struct. Biol.* 2007; 17:633–640. [PubMed: 18024008]
18. Lezon TR, et al. *Handbook on Biological Networks*. 2009:129–158.
19. Wriggers W, Schulten K. *Proteins: Struct., Func., Genetics*. 1997; 29:1–14.
20. Koike R, et al. *J. Mol. Biol.* 2014; 426:752–762. [PubMed: 24189051]
21. Wieninger SA, Ullmann GM. *J. Chem. Theory. Comput.* 2015
22. Wieninger SA, et al. *J. Mol. Biol.* 2011; 409:450–465. [PubMed: 21477597]
23. Hayward S, Berendsen HJ. *Proteins: Struct., Func., Genetics*. 1998; 30:144–154.
24. Cui Q, et al. *J. Mol. Biol.* 2004; 340:345–372. [PubMed: 15201057]
25. Montange RK, Batey RT. *Annu. Rev. Biophys.* 2008; 37:117–133. [PubMed: 18573075]
26. Draper DE, et al. *Annu. Rev. Biophys. Biomol. Struct.* 2005; 34:221–243. [PubMed: 15869389]
27. Denning EJ, et al. *Biophys. Chem.* 2013; 184:8–16. [PubMed: 24012912]
28. Priyakumar UD, et al. *J. Am. Chem. Soc.* 2009; 131:17759–17761. [PubMed: 19919063]
29. Eichhorn CD, et al. *Nucleic Acids Res.* 2011:gkr833.
30. Gong Z, et al. *PloS one*. 2014; 9:e92247. [PubMed: 24663240]
31. Nygaard R, et al. *Cell*. 2013; 152:532–542. [PubMed: 23374348]
32. Huxley H. *Cold Spring Harbor Symposia on Quantitative Biology*. 1973:361–376.
33. McGough A, et al. *J. Cell. Biol.* 1997; 138:771–781. [PubMed: 9265645]
34. Sindelar CV, Downing KH. *Proc. Natl. Acad. Sci. U.S.A.* 2010; 107:4111–4116. [PubMed: 20160108]
35. Rasmussen SG, et al. *Nature*. 2011; 469:175–180. [PubMed: 21228869]
36. Bueren-Calabuig JA, Michel J. *P.L.O.S. Comput. Biol.* 2015; 11:e1004282.
37. Kidd BA, et al. *P.L.O.S. Comput. Biol.* 2009; 5:e1000484.
38. Badireddy S, et al. *Mol. Cell. Proteom.* 2011; 10:M110.004390.
39. Kobilka BK, Deupi X. *Trends Pharmacol. Sci.* 2007; 28:397–406. [PubMed: 17629961]
40. Vaidehi N, Kenakin T. *Curr. Opin. Pharmacol.* 2010; 10:775–781. [PubMed: 20933468]
41. Dror RO, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:4689–4694. [PubMed: 19258456]
42. Dror RO, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:18684–18689. [PubMed: 22031696]
43. Brooks BR, et al. *J. Comput. Chem.* 2009; 30:1545–1614. [PubMed: 19444816]
44. Hess B, et al. *J. Chem. Theory. Comput.* 2008; 4:435–447. [PubMed: 26620784]
45. U'Prichard D, et al. *J. Biol. Chem.* 1978; 253:5090–5102. [PubMed: 209026]
46. Larsson I. *J. Vet. Pharmacol. Therapeut.* 1998; 21:388–392.
47. Innis RB, et al. *Life Sci.* 1979; 24:2255–2264. [PubMed: 41147]
48. Fiser A, et al. *Protein Sci.* 2008; 9:1753–1773. [PubMed: 11045621]
49. Shen, M. y.; Sali, A. *Protein Sci.* 2006; 15:2507–2524. [PubMed: 17075131]
50. Trott O, Olson AJ. *J. Comput. Chem.* 2010; 31:455–461. [PubMed: 19499576]

51. Best RB, et al. *J. Chem. Theory. Comput.* 2012; 8:3257–3273. [PubMed: 23341755]
52. MacKerell AD, et al. *J. Comput. Chem.* 2004; 25:1400–1415. [PubMed: 15185334]
53. Klaua JB, et al. *J. Phys. Chem. B.* 2010; 114:7830–7843. [PubMed: 20496934]
54. Vanommeslaeghe K, et al. *J. Comput. Chem.* 2010; 31:671–690. [PubMed: 19575467]
55. Jorgensen WL, et al. *J Chem Phys.* 1983; 79:926–935.
56. Durell SR, et al. *J. Phys. Chem.* 1994; 98:2198–2202.
57. Jo S, et al. *Biophys. J.* 2009; 97:50–58. [PubMed: 19580743]
58. Levitt M, Lifson S. *J. Mol. Biol.* 1969; 46:269–279. [PubMed: 5360040]
59. Brooks BR, et al. *J. Comput. Chem.* 1983; 4:187–217.
60. Lakkaraju SK, et al. *J. Chem. Info. Model.* 2015; 55:700–708.
61. Feller SE, et al. *J Chem Phys.* 1995; 103:4613–4621.
62. Ryckaert JP, et al. *J. Comput. Phys.* 1977; 23:327–341.
63. Steinbach PJ, Brooks BR. *J. Comput. Chem.* 1994; 15:667–683.
64. Darden T, et al. *J Chem Phys.* 1993; 98:10089.
65. Nosé S. *Mol. Phys.* 1984; 52:255–268.
66. Hoover WG. *Phys. Rev. A.* 1985; 31:1695. [PubMed: 9895674]
67. Parrinello M, Rahman A. *J. Appl. Phys.* 1981; 52:7182–7190.
68. Hess B. *J. Chem. Theory. Comput.* 2008; 4:116–122. [PubMed: 26619985]
69. Hess B, et al. *J. Comput. Chem.* 1997; 18:1463–1472.
70. Miyamoto S, Kollman PA. *J. Comput. Chem.* 1992; 13:952–962.
71. Perera L, et al. *J Chem Phys.* 1995; 102:450–456.
72. Cherezov V, et al. *Science.* 2007; 318:1258–1265. [PubMed: 17962520]
73. Rasmussen SG, et al. *Nature.* 2011; 477:549–555. [PubMed: 21772288]
74. Dror RO, et al. *Science.* 2015; 348:1361–1365. [PubMed: 26089515]
75. Rose AS, et al. *J. Am. Chem. Soc.* 2014; 136:11244–11247. [PubMed: 25046433]
76. Yuan S, et al. *Nat. Commn.* 2014; 5
77. Nygaard R, et al. *Trends Pharmacol. Sci.* 2009; 30:249–259. [PubMed: 19375807]
78. Lee Y, et al. *P.L.O.S. Comput. Biol.* 2015; 11:e1004044–e1004044.
79. Roth CB, et al. *J. Mol. Biol.* 2008; 376:1305–1319. [PubMed: 18222471]
80. Wess J, et al. *EMBO J.* 1993; 12:331. [PubMed: 7679072]
81. Green SA, et al. *Eur J Pharmacol.* 2001; 421:141–147. [PubMed: 11516429]
82. Barak LS, et al. *J. Biol. Chem.* 1994; 269:2790–2795. [PubMed: 7507928]
83. Greasley PJ, et al. *Mol Pharmacol.* 2002; 61:1025–1032. [PubMed: 11961120]
84. van Rhee AM, Jacobson KA. *Drug Devel. Res.* 1996; 37:1. [PubMed: 21921973]
85. Bhattacharya S, Vaidehi N. *Biophys. J.* 2014; 107:422–434. [PubMed: 25028884]
86. Green SA, et al. *Am. J. Resp. Cell. Mol. Biol.* 1995; 13:25–33.
87. Granier S, et al. *J. Biol. Chem.* 2007; 282:13895–13905. [PubMed: 17347144]
88. Seibold A, et al. *Mol Pharmacol.* 2000; 58:1162–1173. [PubMed: 11040066]
89. Kohlhoff KJ, et al. *Nat. Chem.* 2014; 6:15–21. [PubMed: 24345941]
90. Whalen EJ, et al. *Trends Mol. Med.* 2011; 17:126–139. [PubMed: 21183406]
91. Lefkowitz RJ, Caron MG. *Curr. Topic Cell. Regul.* 2014; 28:209–231.

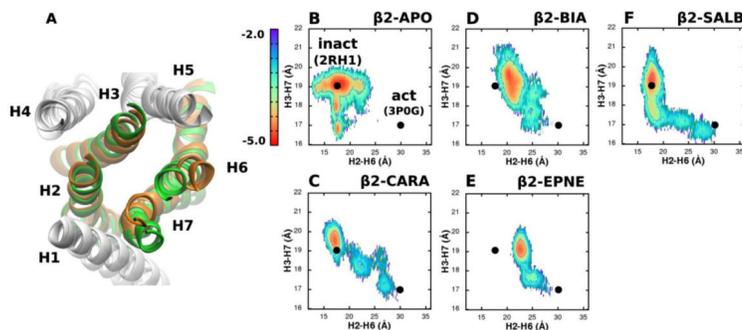


Figure 1.

A) View of the transmembrane helices (H1-H7) from the intracellular side. Structural studies indicate the largest conformational differences are between the inactive (green, PDB: 2RH1) and active (orange, PDB: 3P0G) states across helices H5 and H6, which consequently modulate the cavity between the helices H2-H6 and H3-H7. Approximate free energy surfaces (in kcal/mol) trace the change in the cavity in the B) absence of ligand (APO), presence of C) inverse agonist, carazolol (CARA), agonists D) BI-167107 (BIA), E) epinephrine (EPNE) and F) salbutamol (SALB). H2-H6 and H3-H7 distances in the crystal structures are marked with black dots. Residues used for the calculation were: V67-Y70(H2), S120-V129(H3), C265-H269(H6) and G320-S329(H7).

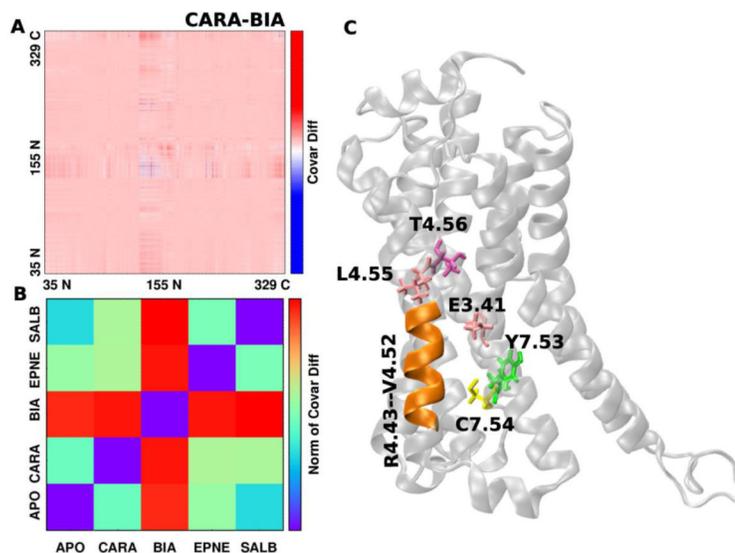


Figure 2.

A) Differences in the covariance matrices of β_2 AR bound to CARA vs. BIA ($\mathbf{D}_{\text{CARA-BIA}}$).
 B) DIRECT-ID analysis quantifies conformational differences across 1- μ s MD trajectories of β_2 AR in the apo state and when bound to CARA, BIA, EPNE and SALB. C) β_2 AR structural elements identified as having biggest changes in conformational dynamics, retrieved from $\mathbf{D}_{\text{CARA-BIA}}$.

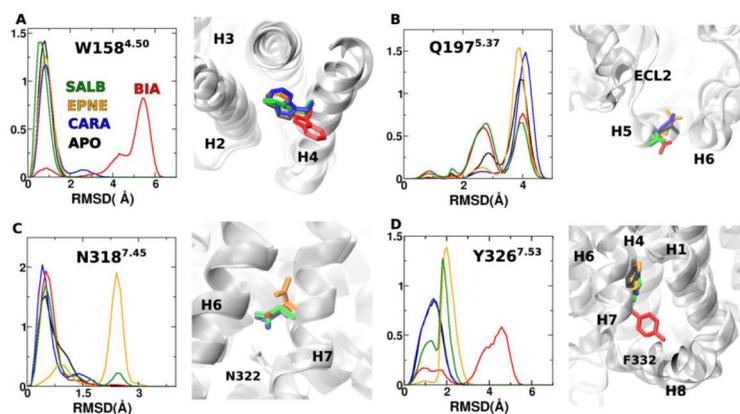


Figure 3. RMSD distributions of A) W158, B) Q197, C) N318 and D) Y326. These residues were identified using DIRECT-ID to have significant differences in their conformational motions between β_2 AR bound to agonists (BIA or EPNE) vs. CARA. Snapshot at 400 ns (where the RMSD values are near the maximum of the probability distributions) reveal the differences in conformations for each of these residues of β_2 AR in the apo state (black), bound to CARA (blue), BIA (red), EPNE (orange) and SALB (green).

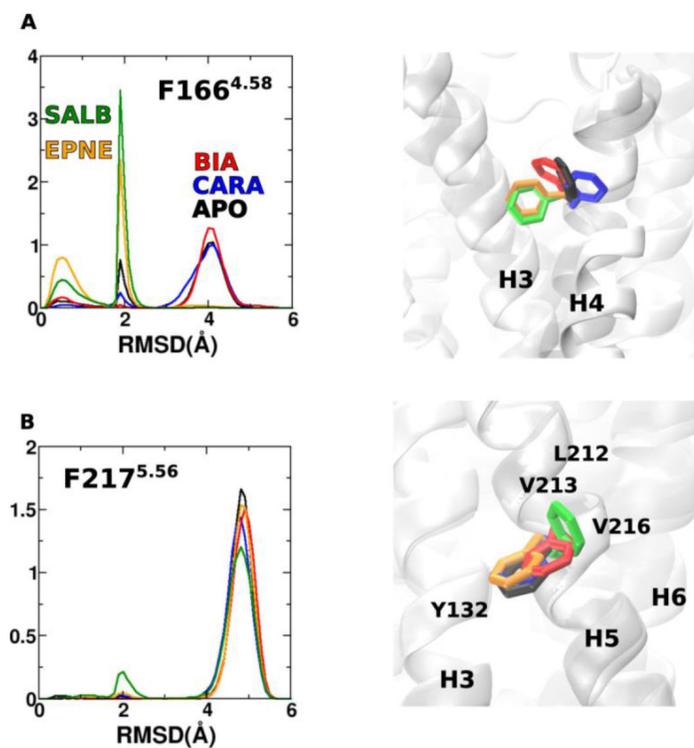


Figure 4. RMSD distributions for A) F166 and B) F217, identified to have large differences in their conformational motion between the SALB-bound β_2 AR and apo state. Snapshots at 400 ns (where the RMSD values are near the maximum of the probability distributions) reveal the differences in conformations for these residues across the different simulations. Coloring scheme is the same as in Fig. 3.

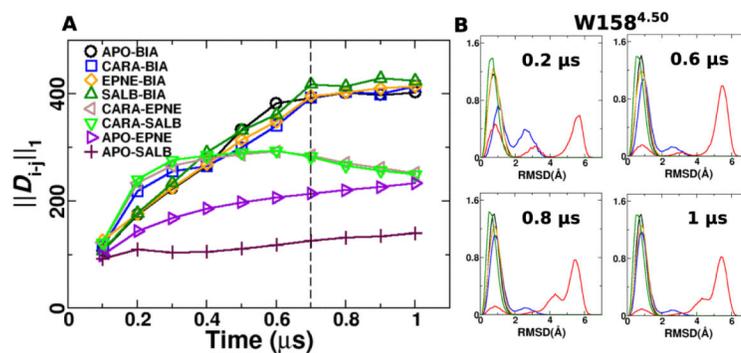


Figure 5.

A) Convergence of $\|D_{i-j}\|_1$ through 1- μ s simulations. Values were calculated in 0.1- μ s increments. B) RMSD distributions of W158 in β_2 AR from apo simulations (black) and when bound to CARA (blue), BIA (red), EPNE (orange) and SALB (green) calculated through 0.2, 0.6, 0.8 and 1 μ s of the MD simulations.

Table 1

β_2 AR residues with the biggest changes in conformational motions between two trajectories. Ballesteros-Weinstein numbering of the residues is marked in the superscript.

Trajectories	$\ D_{i-j}\ _1$	Residues with biggest change in conformational motions	Published mutations and their effect on the receptor
APO-CARA	194.4	R151 ^{4.43} , F166 ^{4.58} , I169 ^{4.61} , Q170 ^{4.62} , Q197 ^{5.37} , L275 ^{6.37} , C327 ^{7.54}	
APO-BIA	403.1	E122 ^{3.41} , R151 ^{4.43} , W158 ^{4.50} , T164 ^{4.56} , Y326 ^{7.53}	E122W: inactivation, ⁷⁹ W158F: reduce agonist affinity, ⁸⁰ T164I: reduce basal activity, ⁸¹ Y326A: inactivation ⁸²
APO-EPNE	233.6	G35 ^{1.34} , M36 ^{1.35} , I38 ^{1.37} , L42 ^{1.41} , F208 ^{5.47} , F217 ^{5.56} , Y219 ^{5.58} , L275 ^{6.37} , I277 ^{6.39} , F282 ^{6.44} , N318 ^{7.45} , Y326 ^{7.53} , R328 ^{7.55}	
APO-SALB	140.2	G35 ^{1.34} , F49 ^{1.48} , F166 ^{4.58} , Q170 ^{4.62} , Q197 ^{5.37} , V206 ^{5.46} , F217 ^{5.56} , I278 ^{6.40} , H296 ^{6.58} , I298 ^{6.60} , E306 ^{7.32} , L310 ^{7.36} , I325 ^{7.52} , R328 ^{7.55}	
CARA-BIA	415.1	E122 ^{3.41} , R151 ^{4.43} , V160 ^{4.52} , L163 ^{4.55} , T164 ^{4.56} , Y326 ^{7.53} , C327 ^{7.54}	T164V: reduce agonist-driven activation ⁸¹
CARA-EPNE	252.4	R151 ^{4.43} , F166 ^{4.58} , I169 ^{4.61} , Q170 ^{4.62} , Q197 ^{5.37} , F217 ^{5.56} , L275 ^{6.37} , E306 ^{7.32} , N318 ^{7.45} , C327 ^{7.54}	L275A: reduce agonist activation ⁸³
CARA-SALB	249.4	R151 ^{4.43} , F166 ^{4.58} , I169 ^{4.61} , Q170 ^{4.62} , Q197 ^{5.37} , F217 ^{5.56} , L275 ^{6.37} , E306 ^{7.32} , C327 ^{7.54}	
EPNE-BIA	414.8	E122 ^{3.41} , R151 ^{4.43} , I153 ^{4.45} , I159 ^{4.51} , L163 ^{4.55} , T164 ^{4.56} , F217 ^{5.56} , N318 ^{7.45} , Y326 ^{7.53}	F217A: reduce activity, ⁸⁴ N318A: inactivation ⁸²
EPNE-SALB	209.4	G35 ^{1.34} , M36 ^{1.35} , I38 ^{1.37} , L42 ^{1.41} , F208 ^{5.47} , F217 ^{5.56} , Y219 ^{5.58} , L275 ^{6.37} , I277 ^{6.39} , T281 ^{6.43} , N318 ^{7.45} , Y326 ^{7.53} , R328 ^{7.55}	
SALB-BIA	425.0	E122 ^{3.41} , R151 ^{4.43} , L155 ^{4.47} , V157 ^{4.49} , I159 ^{4.51} , T164 ^{4.56} , F217 ^{5.56} , Y326 ^{7.53} , C327 ^{7.54}	