Dynamic allosteric networks drive adenosine A_1 receptor activation and G-protein coupling

Miguel A. Maria-Solano[†,*] and Sun Choi[†,*]

[†] Global AI Drug Discovery Center, College of Pharmacy and Graduate School of Pharmaceutical Science, Ewha Womans University, Seoul 03760, Republic of Korea

*Corresponding authors. Emails: miguel.mariasolano@ewha.ac.kr / sunchoi@ewha.ac.kr

ABSTRACT

G-protein coupled receptors (GPCRs) present specific activation pathways and signaling among receptor subtypes. Hence, an extensive knowledge of the structural dynamics of the receptor is critical for the development of therapeutics. Here, we target the adenosine A_1 receptor (A_1R), for which a negligible number of drugs have been approved. We combine molecular dynamics simulations, enhanced sampling techniques, network theory and pocket detection to decipher the activation pathway of A_1R, decode the allosteric networks and identify transient pockets. The A_1R activation pathway reveal hidden intermediate and pre-active states together with the inactive and fully-active states observed experimentally. The protein energy networks computed throughout these conformational states successfully unravel the extra and intracellular allosteric centers and the communication pathways that couples them. We observe that the allosteric networks are dynamic, being increased along activation and fine-tuned in presence of the trimeric G-proteins. Overlap of transient pockets and energy networks uncover how the allosteric coupling between pockets and distinct functional regions of the receptor is altered along activation. By an in-depth analysis of the bridge between activation pathway, energy networks and transient pockets, we provide a further understanding of A_1R. This information can be useful to ease the design of allosteric modulators for A_1R.

INTRODUCTION

Biomolecules present intrinsic dynamism and plasticity to respond to physiological changes.[1] The binding of ligands in the orthosteric site located at the extracellular region can establish dynamic and chemical communication with the intracellular region resulting in key structural rearrangements that trigger a particular response.[2] This communication between distant protein sites is called allostery and its understanding presents a current challenge for many protein complexes.[1, 3-5] Allosteric modulators are capable to bind regions other than orthosteric sites and propagate communication networks to other functional regions of the receptor.[6] A complete information for the structure-based design of allosteric modulators involves not only the exploration of the receptor activation pathway and the identification of binding sites,[7] but also the characterization of the allosteric pathways, which is a complicated task. However, allosteric drugs offer the advantage of higher selectivity respect to conventional drugs due to the greater sequence variability of allosteric sites.[8, 9]

In G-protein coupled receptors (GPCRs), the binding of endogenous agonist in the orthosteric site reshapes the conformational ensemble of a Transmembrane (TM) helix to prime the binding and activation of G-proteins.[10-12] The conformational landscapes of GPCRs associated with the receptor activation are complex, with some general principles in common and also receptor-specific differences.[13, 14] For many GPCRs it has been observed that (1) inactive and pre-active states are in dynamic equilibrium in the apo form, (2) the action of the agonist induce subtle population shifts involving the stabilization of intermediate and pre-active states and/or decreasing their rates of interconversion and (3) the combined action of both, agonist and G-protein binding stabilizes a more rigid fully-active state.[11, 13, 15-17] The complexity of GPCRs activation landscapes difficult its experimental and computational characterization. However, the study of its system-specific properties opens the door for the design of allosteric drugs.

The adenosine A_1 receptor (A_1R) is a member of the class A G protein-coupled receptor (GPCR) family that preferentially couples with Gi/o proteins. It is widely distributed in multiple organs mediating a variety of physiological processes, including those in the brain and the heart. Thus,

A_1R has significant therapeutic potential in the treatment of numerous diseases and disorders.[18] In fact, it has been targeted for the pain management through allosteric modulation although without success in clinical trials.[19, 20] Fortunately, X-ray crystallography and Cryo-electron microscopy (Cryo-EM) captured A_1R in an inactive and active states revealing a notorious inward-to-outward conformational transition of TM6 (**Figure 1A** and **S1**).[19, 21, 22] The most recently released active state structure is resolved with adenosine, trimeric G-protein and also with a positive allosteric modulator (PAM) located in an extrahelical region.[19] Computational and mutagenesis studies have reported interesting insights about the A_1R activation and allosteric modulation identifying important residues for the signaling efficacy of agonists and cooperativity of PAMs.[19, 23, 24] Despite all these achievements, a detailed characterization of the allosteric networks that drive receptor activation and G-proteins binding is still missing.

In this work, we develop a computational workflow tailored to decipher the interplay between the receptor activation pathway, the allosteric communication networks and transient pockets. First, we used conventional molecular dynamics (MD) simulations coupled with enhanced sampling techniques to reconstruct the receptor activation conformational landscape of the inward-to-outward TM6 transition revealing the inactive, intermediate, pre-active and fully-active conformational states. Second, we study the dynamic communication energy networks throughout the conformational states sampled successfully capturing the extra and intracellular communicating centers and the pathways that interconnect them. We observe that the allosteric communication is enhanced along the receptor activation and fine-tuned in presence of the trimeric G-protein. Third, we used a geometric-based approach to search for the formation transient pockets. By studying the connection between these three elements we give a complete dynamic picture of the A_1R activation that is essential for the design of specific allosteric modulators. This *in silico* approach can be also applied to uncover the activation landscape, allosteric networks and transient pockets of related GPCRs, which is of interest for allosteric drug design.

RESULTS

The free energy landscape of A_1R activation reveals intermediate and pre-active states. To uncover the activation pathway of A_1R in presence of its endogenous agonist adenosine (A_1R-ADO), we reconstructed the free energy landscape (FEL) of A_1R-ADO associated with the TM6 inward-to-outward transition observed by X-ray and Cryo-EM data (**Figure 1A**). Specifically, we focus on two features: the TM6 torsion and the center of mass (COM) distances between the TM3 and TM6 intracellular ends (**Figure 1B**). Initially, we performed conventional molecular dynamics (cMD) simulations starting from both, the inactive (PDB 5N2S) and the active (PDB 6D9H) structures in order to increase the sampling of the activation conformational space. For each starting point we computed 3 replicas of 500ns, which is a reasonable simulation time to provide an initial sampling of the receptor activation. Either starting from the inactive or active structures the complete TM6 Inward-to-Outward transition is not sampled (**Figure 1C**). However, when starting from the active structure, an intermediate state centered ca. at TM6 torsion (-140°) and TM3-TM6 distance (15 Å) coordinates is substantially populated. This intermediate structure still presents a rather high degree of TM6 torsion while the TM3-TM6 distances are more shortened. In contrast, when starting from the inactive structure, the receptor mostly samples inactive conformations exhibiting a low probability to progress towards the intermediate state, thus suggesting a higher inactive-to-intermediate transition time scale. For a better characterization of the conformational landscape of the A_1R-ADO activation and check the importance of the intermediate state, we relied on metadynamics simulations. The primary objective of this calculation is to identify and reconstruct the major conformational states that are involved in the pathway of receptor activation. Metadynamics is a powerful method that has been successfully used to study complex conformational transitions in proteins,[25-27] including GPCRs.[12, 15] In particular, we performed well-tempered metadynamics (WT-MetaD) simulations using the walkers approach (see Materials and methods). We selected 10 representative structures along the activation pathway sampled in the cMD as starting points for the walker replicas (see **Figure S2**). After 250 ns of accumulated time, we successfully reconstructed the major conformational states of the FEL (see convergence assessment in **Figure S3 and S4**).

The FEL in presence of adenosine confirms that A_1R presents three major states in dynamic equilibrium (inactive, intermediate and pre-active) while the fully-active Cryo-EM like state is not significantly populated. The relative stabilities of the three states shows that the most stable state is the inactive, followed by the pre-active and intermediate states, which are slightly higher in energy. The inactive state resembles the inactive X-ray coordinates showing short TM3-TM6 distances and low TM6 torsion values (state 1 in **Figure 1D-E**). The inactive conformations are stabilized by a tight energy coupling between TM3 and TM6 (see below). As the activation progresses, TM6 torsion evolves further than the TM3-TM6 distances. At this point, the remaining energy coupling between TM3 and TM6 ends hampers the progression of the TM6 outward transition. This tug of war between forces provokes the adoption of a torsion in the TM6 end that is signature of the intermediate conformations (state 2 in **Figure 1D-E**). This notorious torsion evolves to a more stable intermediate conformation (state 3 in **Figure 1D-E**). Finally, a pre-active local energy minimum is reached completing the TM6 transition, which involves the complete break between TM6 and TM3 energy coupling (see below). The pre-active state presents similar TM3-TM6 distances to the fully-active Cryo-EM structure. However, it exhibits higher TM6 torsion values (state 4 in **Figure 1D-E**). In addition, an extra opening of TM6 is accessible (state 5 in **Figure 1D-E**). All together suggest that the coupling of the G-proteins is required to stabilize the adoption of the fully-active Cryo-EM like structure (see next sections). As a complementary analysis, we conducted the reweighting of the metadynamics simulations[28] to determine the free energy as a function of previously identified A_1R micro-switches (ionic-lock, PIF motif, water-lock and toggle switch). The fact that we capture the distinct energy barriers associated with unbiased micro-switches highlights the accuracy of the metadynamics simulations in reproducing the pathway of activation and provides useful information to guide the selection of collective variables for future GPCR landscape calculations (**Figure S5, S6 and S7**). Once deciphered the A_1R-ADO FEL of activation, our study follows on investigating the receptor-specific allosteric properties that harbor these inactive, intermediate and pre-active states.

Energy Networks captures the dynamic allosteric pathways along A_1R activation. A complete understanding of allosteric modulation involves the decoding of the communication pathways that dynamically couples distinct protein sites. Despite the difficulties, network theory has been successfully applied to uncover the allosteric communication pathways in protein complexes[25, 29], including GPCRs.[30, 31] In order to trace down the allosteric pathways in A_1R-ADO, we relied on the protein energy networks (PEN) approach.[32] First, we use the get Residue Interaction eNergies and Networks (gRINN)[32] tool to calculate the pairwise residue interaction energies along the A_1R-ADO conformational ensemble and obtain the mean interaction energy matrix (IEM). Second, we processed the mean IEM into the shortest-path map (SPM)[29, 33, 34] tool to construct and visualize the PEN graph.
The analysis of the PEN associated with the A_1R-ADO ensemble shows that the extracellular region can communicate with the intracellular region through multiple energy pathways (see **Figure 2A**). In the extracellular region, ECL2 and ECL3 presents a center of communication connecting with TM4, TM5, TM1 and TM6. Energy pathways involving TM5, TM6 and TM4 propagates from the extracellular ECL2/ECL3 regions and link with TM3 in the intracellular region. In addition, TM3 establish connections with TM2 and TM7 that communicates back to the ECL2 by crosslinking with TM1 (**Figure 2A-B**). TM6 is then energetically coupled with the extracellular region through ECL3/ECL2 and with the intracellular region through TM3. In fact, the TM3 end plays a central role for the allosteric communication in the intracellular region. Specifically, $R108^{3.53}$ is a communication hub forming and hydrogen networking with $D229^{6.30}$, $D104^{3.49}$ (ionic-lock micro-switch) and $D326^{C\text{-}Ter}$. Regarding the energy communication at the orthosteric site, adenosine samples a wide range of poses in its large binding site[19, 35] performing transient interactions with many residues of the PEN, thus establishing multiple transient energy communication that propagates towards the TM6 intracellular end (see **Figure S8**).

In order to provide further insights in the allosteric communication along the activation pathway we decided to split the analysis into the inactive, intermediate and pre-active states (Figure **3A** and **S9**). We started the analysis with the inactive PEN, which shows that the extracellular ECL2 center is only connected to the intracellular region through TM2 and TM1. In the intracellular region, $R108^{3.53}$ establish a tight communication with $D104^{3.49}$, $D229^{6.30}$ and $D326^{C\text{-}Ter}$, see **Figure 3A-B**. Note that $D229^{6.30}$ can also communicate with $R105^{3.50}$. This hydrogen-bonded network that captures the ionic-lock micro-switch is signature of the inactive in-ward conformation of TM6. In fact, $D229^{6.30}$-$R108^{3.53}$ interaction is pivotal to attain the inactive conformation of the receptor.

Interestingly D326$^{C\text{-}Ter}$ plays an important role in the energy communication. The flexible C-Ter region can partially occupy the G-proteins binding site and perform stable communication with R108$^{3.53}$. We follow the analysis with the PEN of the intermediate ensemble. In this case, the extracellular communication center is reduced and ECL2/ECL3 region is connected to the intracellular region through TM7 and TM1. Interestingly, the communication through TM6 start to take place partially in the intracellular region. In contrast with the inactive ensemble, D229$^{6.30}$ now mainly communicates with R105$^{3.50}$ instead of R108$^{3.53}$ in the intracellular region, see **Figure 3A-B**. Therefore, at this point the dynamics of the ionic-lock is altered. This transient interaction is key to prevent TM6 opening after the tension generated in the last segment of TM6 and provokes a torsion in the TM6 intracellular end that captures the receptor in an intermediate state (described above). In this scenario, R108$^{3.53}$ communicates with D42$^{2.37}$ and D104$^{3.49}$. Note that the C-Ter does not communicate with TM3, which explains the transience of this interaction due to the C-Ter flexibility. Finally, in the pre-active ensemble, the extracellular ECL2 networks are recovered and connect to the intrahelical region through multiple pathways, including TM2, TM4, TM5, TM6 and TM7. Interestingly, the PEN capture Y200$^{5.58}$ and Y288$^{7.53}$ in the TM5 and TM7 pathways, respectively. These tyrosine residues have been found to stabilize the active state through a hydrogen bond coordinated by a bridging water molecule in the so-called water-lock. However, the water-lock bridge is not observed because non-protein molecules are excluded in the PEN calculation, which points out a major limitation of the methodology. Regarding the TM6 pathway, its partial communication observed in the intermediate state progresses further capturing the toggle switch (W247$^{6.48}$) and it is completed linking with TM5 (**Figure 3A**). This enhanced communication between the intra- and extracellular region is driven by the complete opening of TM6. As expected, the TM6 opening provokes the break of the ionic-lock disconnecting TM6 (D229$^{6.30}$) from TM3 (R108$^{3.53}$ and R105$^{3.50}$). Indeed, the R108$^{3.53}$ node loses prominence in the intracellular region. However, R108$^{3.53}$ still can communicate with D326$^{C\text{-}Ter}$, which may compensate the break of TM3-TM6 ends (**Figure 3A-B**). Complementary insights are gained by computing the histograms of relevant micro-switches along the inactive, intermediate and pre-active states (**Figure S10**).

At this point, we wonder if the positions captured in the PEN coincide with allosteric residues previously identified in mutagenesis studies that affect the allosteric responses of the receptor. For the ECL2/ECL3 extracellular region 9 allosteric residues have been identified that affect the efficacy of orthosteric agonist.[23, 36] Among all of them, up to 7 are captured in the ECL2/ECL3 PEN communication center (**W156$^{45.37}$, N148$^{45.29}$, K173$^{45.54}$, K265^{67}, E170$^{45.51}$, S150$^{45.31}$** and **T270$^{7.35}$**) and the other 3 are located in the ECL2 alpha helix that connects with the PEN through S150. In addition, other 4 allosteric residues located in a TM1-7 extrahelical region have been recently reported to be important for PAM cooperativity.[19] The PEN directly captures **S246$^{6.47}$**, **L242$^{6.43}$** and **L245$^{6.46}$**, and the other residue (**G279$^{7.44}$**) is located adjacent to the PEN residue H278 and S246. Note that the experimentally identified allosteric residues captured in the PEN are labeled with a red asterisk in **Figure 2 A-C**. It is worth mentioning that among the allosteric residues captured in the PEN most of them are identified in the pre-active ensemble while only some of them are identified in the inactive and intermediate states. This is due to the enhanced allosteric communication observed upon activation. In summary, the PEN analysis along the conformational states captures many important positions for the allosteric mechanisms of the receptor and provides a dynamic view of the protein energy communication along the activation of the receptor. The TM6 in-ward to outward transition reveals that the TM6 allosteric pathway along the receptor is missing in the inactive state, partially formed in the intermediate state and completely established upon the receptor activation. Indeed, the pre-active state is characterized by an increased allosteric communication though multiple pathways.

G-Proteins binding stabilizes the fully-active state and fine-tune the allosteric communication. The next unknown we were intrigued to investigate is how the binding of G-proteins affects the A_1R conformational ensemble and the allosteric networks. To that end, we performed 3 cMD replicas of 500 ns of A_1R in presence of both, adenosine and heterotrimeric G_{i2} protein (referred as ADO-A_1R-G_{i2} complex). As previously, the MD data was plotted as a function of the TM6 torsion and TM3-TM6 ends distances (**Figure 4A**). The population analysis of the receptor activation shows that the presence of the slim $G\alpha_{i2}$-α5 helix in the A_1R intracellular cavity prevents TM6 to sample intermediate and inactive conformations. Indeed, A_1R-G_{i2} displays restrictive conformational dynamics by only sampling one major conformational state. This conformational state overlaps with both, the pre-active state and the fully-active state

corresponding to the active Cryo-EM structure (**Figure 4B-C**). Thus, the binding of the G-proteins induces a population shift in A_1R towards fully-active conformations, as described previously in other GPCRs.[11-14]

To study the allosteric pathways of the ADO-A_1R-G_{i2} complex, we computed the PEN considering the intra (A_1R-A_1R) and inter (A_1R-$G\alpha_{i2}$) interactions. The PEN analysis shows some similarities respect to the PEN of the pre-active ensemble (**Figure 5**). In both cases, TM6 allosteric communication is completed and TM7 presents an important communication pathway connecting the intra- and extracellular regions. However, the presence of G-Proteins finely-tunes the PEN in some aspects: (1) The ECL2 extracellular communication center propagates toward the ECL2 helix capturing the **W156$^{45.37}$** allosteric residue that was not identified in previous PEN analysis. (2) The communication pathways between extra and intracellular regions are refined only passing through TM2, TM1, TM6 and TM7 and not using TM4 and TM5. Certainly, TM7 becomes the major communication pathway and another allosteric residue **G279$^{7.44}$** is captured as an important communication node. (3) As expected, the C-Ter does not communicate with TM3 because the end of the α5 helix of $G\alpha_{i2}$ ($G\alpha_{i2}$-α5 helix) replaces the C-Ter communication position. As a consequence, TM3 (R108) communicates with TM2 (D42) and most importantly with $G\alpha_{i2}$-α5 helix (D667). The latest becomes a communication hub in the intracellular region by also connecting with H8 (K294). The inclusion of $G\alpha_{i2}$ also causes that TM6 (E229) communicates with TM5 (R208), **Figure 5**. Thus, the receptor signaling attains a specific profile of protein energy communication networks triggered by the $G\alpha_{i2}$ protein coupling.

Identification of transient pockets as potential allosteric sites. The identification of transient pockets formed along specific receptor activation profiles is useful to guide the design of allosteric drugs. We use a geometry-based algorithm (MDpocket)[37] in order to search for transient pockets in the inactive, intermediate, pre-active and fully-active ensembles of A_1R. The MDpocket output provides a normalized frequency map that allows the visualization of the frequency of a pocket formation along each conformational ensemble studied.

We identified multiple pockets in the upper and lower regions of the receptor. Interestingly, we observe that pockets can change its shape and frequency of formation along the different conformational states (**Figure 6A**). Focusing in the upper region pockets (**PA-D**), **PA** is present the ECL2 vestibule with low frequency. This region has been proposed to bind A_1R PAMs on the basis of computational and mutagenesis studies.[23, 35, 36] **PB** is located at the adenosine binding site and extends to the adjacent secondary pocked observed by X-ray data in the inactive A_1R structure.[22] Due to the **PB** large size, it overlaps with many allosteric modulators observed in class A (PDBs 4MQT,[38] 5NDD,[39] 4MBS[40]) and class C (PDBs 5CGC,[41] 5CGD,[41] 6FFH[42] and 6FFI[42]) GPCR structures. **PE** corresponds to the protein internal channel and overlaps with the NAM sodium Ion site (PDB 4N6H),[43] which is narrowed along activation (**Figure 6A**). **PD** corresponds to the MIPS521 PAM shallow pocket exposed to the lipid-detergent interface (PDB 7LD3).[19] Up to date, this is the only allosteric modulator which structure has been resolved in complex with A_1R. Given its relevance for this study, PB is explored in more detail in the next section. We also identify a pocket (**PC**) located in an extrahelical TM1 and TM7 region. It matches with the polar region of the Monooleoylglycerol molecule captured in a GPCR structure (PDB 4MBS) 35. **PC** is energetically coupled with TM1 and TM7 PEN residues. For a complete description all pockets location and containing PEN residues, see **Table S1 and S2**.

Regarding the lower part of the receptor, both the outer surface (extrahelical) and the inner surface (intrahelical) presents many cavities. Some of them are open along the activation pathway (**PF**, **PL**) and others are more predominant (**PJ, PH**) or only formed (**PM**, **PK**, **PN**) in some states. **PF** is formed in all states and matches with the AP8 allosteric modulator site found in the free fatty acid receptor 1 (PDB 5TZY)[44] (**Figure 6A**) and also with cholesterol and other lipid molecules found in other GPCR structures (PDB 5LWE,[45] 4PHU,[46] 5TZR,[44] and 4XNV[47]). Among GPCR allosteric modulators co-crystalized in the intracellular surface, only Vercircon (5LWE)[45] overlaps partially with **PL, Figure 6A**. This suggest that the shape and location of the A_1R pockets located in the lower part is highly system-specific. In fact, the shape of the lower region changes substantially along activation. **PG** and **PF** pockets are formed from the intermediate to the fully-active state and are coupled to TM5-TM6 and TM2-TM3-TM4 PEN residues, respectively. **PM** supposes an interesting pocket because it is only predominant in the intermediate state. In addition, it contains many PEN residues of the extrahelical TM1-TM2-TM4

region. **PK** is only formed in the intermediate and pre-active states and contains PEN residues of the extrahelical TM1-H8 region, while **PI** is only formed in the pre-active and fully-active states and is energetically coupled with TM6 and TM7. **PH** and **PN** are located in the extrahelical TM5-TM6 region. **PH** is more predominant in the pre-active and fully-active state containing many PEN residues while **PN** is only formed in the pre-active and fully active states but it is less energetically connected (**Figure 6B**). The overlap between pockets and PEN provides a view of how the distinct pockets are allosterically coupled between them and with other functional regions of the receptor (**Figure 6B** and **Table S1-2**).

ADO and MIPS51 PAM have a significant impact on the energy networks. In order to establish a connection between the energy networks and the mode of action of allosteric modulators, we focus on exploring the effect of MIPS521 positive allosteric modulator (PAM) and ADO agonist as a proof of concept. Experimental assays and Gaussian accelerated MD determined that MIPS521 PAM increases the binding affinity of ADO in the orthosteric site.[19] Thus, **PB** and **PD** must be allosterically coupled. Among MIPS521 PAM pocket residues, only **L242$^{6.43}$**, **L245$^{6.46}$**, **S246$^{6.47}$** and **G279$^{7.44}$** were experimentally found to affect the PAM cooperativity (**Figure S11**). Interestingly, the PEN obtained in presence of ADO captures these key residues along activation, including TM6 (**L242$^{6.43}$** and **L245$^{6.46}$**) in the intermediate, **L242$^{6.43}$** and **S246$^{6.47}$** in the pre-active and **L242$^{6.43}$** and TM7(**G279$^{7.44}$**) allosteric residues in the fully-active ensemble (**Figure 7**). Indeed, **G279$^{7.44}$** becomes a key node in the PEN of the fully-active ensemble. This evidence suggests that although both PD and PB are open in all conformational states, their energy coupling is particularly stronger during the receptor activation.

This prompted us to investigate whether the binding of ADO and the MIPS521 PAM can affect the allosteric communication between PB and PD sites. To that end, we performed cMD of the heterotrimeric G_{i2} protein ADO-A_1R-G_{i2} complex in presence of the PAM (PAM-ADO-A_1R-G_{i2} complex) and in absence of adenosine (A_1R-G_{i2} complex) in order to compute their conformational landscape and energy networks following the same protocol for the ADO-A_1R-G_{i2} complex (**Figure 8 and S12**). The analysis of the PEN of A_1R-G_{i2} complex reveals that in the absence ADO, the receptor displays a reduced allosteric communication between PB and functional regions of the receptor, such as the extracellular allosteric center, TM6 and PD allosteric site. As expected, the presence of ADO restores the allosteric coupling between PB and TM6, which could explain the increase of receptor activity associated with agonist binding. Additionally, our analysis of the PAM-ADO-A_1R-G_{i2} complex shows that the PAM reinforces the TM7-ECL3-ECL2 allosteric pathway that couple PD with PB, and ECL2 now communicates to the intracellular region through TM5 (**Figure 8**). Notably, a recently published study reported that the orthosteric pocket contracts after ADO binding, as demonstrated by shortened distances of the so-called vestibular lid (defined as the sum of length of the triangle perimeters formed by E170$^{45.51}$-Y271$^{7.36}$-E172$^{45.53}$ interacting residues) and the E172$^{45.53}$-K265^{67} salt bridge.[48] Remarkably, the TM7-ECL3-ECL2 enhanced pathway by PAM effect contains the vestibular lid and the E172$^{45.53}$-K265^{67} salt bridge residues (**Figure 8**). This suggest that PAM promotes the contraction of PB, leading to the stabilization of the ADO-bound state. Thus, the enhanced energy coupling between PB and PD may be responsible for the increase in the binding affinity of ADO in presence of the PAM, as observed experimentally.[19] This data indicates that allosteric modulators are able to enhance and redistribute the energy networks, which is likely attributed for their effects on the receptor activity.

DISCUSSION

A comprehensive knowledge of the allosteric properties of receptors is essential for the design of allosteric drugs. It involves the deciphering of activation conformational landscapes, the decoding of allosteric networks and the characterization of transient pockets that are allosterically coupled with functional regions of the receptor. Here, we focus on the A_1R due to the failure in the development of efficient and safe allosteric modulators reported up to date.[18, 19, 23]

The A_1R activation FEL associated with the TM6 inward-to-outward transition reveals that the receptor is in dynamic equilibrium with inactive, intermediate and pre-active states, where the inactive is the most stable and the intermediate and pre-active states are separated by a lower energy barrier. This fast-conformational transition observed in presence of adenosine may favor the binding and activation of G-proteins.[12, 13] Regarding the activation pathway, the inactive state of TM6 that resembles the inactive X-ray structure progresses towards an intermediate state

that is characterized by a rather high TM6 torsion and a modest TM6 opening due to the tight TM3-TM6 energy coupling in the intracellular ends. This results in a torsion in the end of TM6, which is signature of the intermediate conformations. This torsion evolves to a more stable intermediate state that exhibit higher TM3-TM6 distances. Finally, TM6 progresses towards a complete opening reaching the pre-active state, in which the fully-active Cryo-EM structure is not highly populated. Interestingly a larger outward movement of the TM6 is accessible, as captured in other class A GPCR receptors.[49, 50] Upon G-proteins coupling, the slim $G\alpha_{i2}$-α5 helix induces a population shift towards TM6 fully-active conformations that resembles the active Cryo-EM structure, and the intermediate and inactive states are not accessible. This data is in line with the combined activation mechanisms described in many GPCRs by means of NMR and computational studies, which states that the combined action of both, agonist and G-proteins is essential to stabilize the adoption of a less dynamic fully-active state.[11, 13, 15, 16]

The conformational progression obtained from the inactive to the adoption of fully-active conformations are relevant to target A_1R drug specificity, specially the hidden intermediate and pre-active states. These conformational ensembles are also valuable to compute the allosteric networks of the receptor along the activation pathway. Such analysis provides a dynamic view of how the allosteric communication evolves along activation. The analysis of the interaction protein energy networks (PEN) captured the extra and intracellular communication centers together with the allosteric pathways that interconnect them.

Focusing on the intracellular region, a tight allosteric commutation is observed in the inactive ensemble between $D104^{3.49}$, $R108^{3.53}$, $D229^{6.30}$ and $D326^{C\text{-}Ter}$, which is key to attain the TM6 in the inactive state. The strong TM6-TM3 hydrogen network weakens in the intermediate state and $D229^{6.30}$ now communicates with $R105^{3.50}$. This transient communication prevents the TM6 opening and captures the receptor in intermediate conformations, that are characterized by a structural torsion in the TM6 end. In the pre-active ensemble TM6 opening provokes the complete loss of TM3-TM6 communication. Interestingly, the flexible $D326^{C\text{-}Ter}$ performs transient communications along the receptor activation with $R108^{3.53}$ and $K116^{4.34}$. These interactions could compensate the TM6-TM3 break upon activation. However, the $D326^{C\text{-}Ter}$ communication takes place in the region where the G-proteins bind. Thus, the binding of $G\alpha_{i2}$-α5 helix must also compensate for the release of the C-ter from the G-proteins binding site. In fact, $G\alpha_{i2}$-α5 helix (D340) replaces the $D326^{C\text{-}Ter}$ communication with $R108^{3.53}$ becoming a hub node in the intracellular allosteric center. Regarding the allosteric pathways, we observe an enhanced communication between the intra- and extracellular centers upon activation. In contrast with the inactive and intermediate states, in the pre-active state multiple energy pathways arises including TM2, TM4, TM5, TM6 and TM7. This is evidenced by the lack of TM6 communication in the inactive state, the partial TM6 communication in the intermediate state and the complete TM6 communication in the pre-active state. Interestingly, the presence of the G-proteins fine-tunes the allosteric communication profile only passing through TM2, TM1, TM6 and TM7. These insights may be interesting for future biased agonism studies. We hypothesize that the binding of drugs that potentially favor the establishment of specific PEN profiles may result in the discrimination between different intracellular partners by preferentially communicating with one of them. Thus, avoiding side effects by activating desired pathways.

Finally, we studied the connection between transient pockets and PEN along the receptor activation in order to unravel the allosteric coupling between pockets and distinct functional regions of the receptor. As a proof of concept, we focus on the PD, which corresponds to the binding site of MIPS52, a positive allosteric modulator (PAM) that increases Adenosine binding affinity in the orthosteric site (PB). Although PD is open in all conformational states the communication between PB and PD is enhanced along activation capturing the allosteric residues that were found to affect its PAM from the intermediate to the fully active. Based on this observation, we hypothesize that drugs that bind pockets and interact with PEN residues, which progresses towards regions of the receptor where function can be altered, may potentially affect the receptor activity through allosteric effects. Additionally, the pocket where the drug binds must be open at least in the conformation state that is targeted. As a practical aspect, virtual screening campaigns could use this information during the design procedure by selecting drug candidates that perform stronger interactions with the PEN contained in the pockets.

To further support this hypothesis, we explored the allosteric effects of ADO and MIPS52 PAM on the PEN. Interestingly, we observed that ADO is crucial for the formation of the extracellular center and its connection with TM6 pathway. Furthermore, MIPS52 PAM reinforces the pathway that connects PB and PD pockets and redistribute other connections. These alterations in the PEN can be related with their mode of action. ADO may increase the activity of the receptor through its communication with TM6 and the PAM may increase ADO binding affinity though stronger energy coupling between PD and PB pockets. These findings imply that the mode of action of allosteric drugs could be predicted depending on how they redistribute the PEN.

CONCLUSIONS

With the combination of free energy landscape construction, the decoding of allosteric networks and the calculation of transient pockets we successfully capture hidden conformational states, the allosteric communication centers/pathways and the transient pockets formed along the A_1R activation. Most importantly, by an in-depth study of the connection between these three elements we provide a complete dynamic view of the A_1R activation. Specifically, we observe that the allosteric communication is progressively enhanced between the extra and intra allosteric centers throughout the inactive, intermediate and pre-active states to be fine-tuned upon the binding of G proteins in the fully-active state. In fact, not only the allosteric networks are dynamic but also the shape and frequency of formation of the transient pockets change along the different conformational states. Overlap of the energy networks and transient pockets uncover how the allosteric coupling between pockets and distinct regions of the receptor is altered along the receptor activation pathway. As a proof of concept, Adenosine and a previously experimentally determined positive allosteric modulator were found to enhance and redistribute the energy networks of the receptor in a manner that is consistent with their respective functions. The prediction of drug effects depending on how they redistribute the protein energy networks presents a promising avenue for drug discovery. All these system-specific structural dynamics understanding provides useful information to advance the design of A_1R allosteric modulators on the basis of structure-based drug design. This computational approach can be also transferable to other GPCRs and related receptors, which is of interest for the design of novel allosteric drugs.

MATERIALS AND METHODS

Conventional Molecular Dynamics (cMD) simulations

System preparation: In total, five systems were prepared: the inactive conformation of A_1R in complex with adenosine (A_1R-ADO, inactive), the active conformation in complex with Adenosine (A_1R-ADO, active), the active conformation in complex with adenosine and heterotrimeric G_{i2} proteins (ADO-A_1R-G_{i2}, active), the active conformation in complex with the heterotrimeric G_{i2} proteins (A_1R-G_{i2}, active) and the active conformation in complex with adenosine, the MIPS521 positive allosteric modulator and heterotrimeric G_{i2} proteins (PAM-ADO-A_1R-G_{i2}, active). For the A_1R-ADO inactive system, we used the inactive structure of A_1R in complex with PSB36 (PDB 5N2S)[51] as starting structure. PSB36 was manually removed and adenosine was placed in the orthosteric site by structural alignment with the active Cryo-EM structure of A_1R (PDB 6D9H).[21] The A_1R-ADO active system was obtained by manually removing the G-proteins from PDB 6D9H. Regarding the multimeric complexes, ADO-A_1R-G_{i2} active system was obtained from PDB 6D9H. The crystal structure (PDB 2ODE)[52] was used to reconstruct the missing G_{i2} protein regions and also to place Guanosine-5'-Diphosphate (GDP) and the magnesium ion (Mg^{+2}) by structural alignment. The A_1R-G_{i2} system was generated by manually removing Adenosine from ADO-A_1R-G_{i2} and PAM-ADO-A_1R-G_{i2} was generated by adding the MIPS521 positive allosteric modulator to ADO-A_1R-G_{i2} by structural alignment with PDB 7LD3.The Modeller software[53] was used to reconstruct the missing X-Ray and Cryo-EM regions. MD parameters for the ligands (ADO, PAM and GDP) were generated with the parmchk module of AMBER20[54] using the general amber force-field (GAFF). The atomic charges (RESP model) were obtained using the antechamber module of AMBER20, with partial charges set to fit the electrostatic potential generated at HF/6-31G* level of theory using Gaussian 09.[55] Internal water molecules highly conserved among GPCRs were incorporated into the A_1R internal channel of the inactive and active systems using the HomolWat web server tool. Specifically, HomolWat identified a total of 76 and 100 internal

water molecules that fit into our active and inactive structures, respectively. These water molecules were incorporated from multiple PBD data.[56] All systems were filled into a simulation cell composed by a phosphatidylcholine (POPC) membrane solvated at NaCl 0.15 nM using the packmol memgen tool implemented in AMBER20. The force-fields selected to describe the different molecule types for the MD simulations were ff14SB (protein), GAFF2 (ligands), Lipid17 (membrane) and TIP3P (waters). In addition, two disulfide bonds were created between C80-C169 and C260-C263 residues in A_1R.

Molecular dynamics protocol: The systems were minimized in a two-stage geometry optimization approach. In the first stage, a short minimization of the water molecules positions, with positional restrains on the protein, ligand and P31 atoms of the membrane was performed with a force constant of 10 kcal mol^{-1} $Å^{-2}$ at constant volume periodic boundaries conditions. In the second stage, an unrestrained minimization including all atoms in the simulation cell was carried out. The minimized systems were gently heated in two phases. In the first phase, the temperature was increased from 0K to 100K in a 40 ps step. Harmonic restrains of 10 kcal mol^{-1} $Å^{-2}$ were applied to the protein, ligand and membrane. In the second phase, the temperature was slowly increased from 100K to the production temperature (303.15K) in a 120 ps step. In this case, harmonic restrains of 10 kcal mol^{-1} $Å^{-2}$ were applied to the protein, ligand and P31 atoms of the membrane. The Langevin thermostat was used to control and equalize the temperature. During the heating process the initial velocities were randomized. For the heating and following steps, bonds involving hydrogen were constrained with the SHAKE algorithm and the time step was set at 2 fs, allowing potential inhomogeneities to self-adjust. The equilibration step was performed in three stages. In the first stage, an MD simulation of 5 ns under NVT ensemble and periodic boundary conditions was performed to relax the simulation temperature. In a second stage, an MD simulation of 5 ns under NPT ensemble at a simulation pressure of 1.0 bar was performed to relax the density of the system. The semi-isotropic pressure scaling using the Monte Carlo barostat was selected to control the simulation pressure. In a third stage, an additional MD simulation of 10 ns was performed to further relax the system. After the systems were equilibrated in the NPT ensemble, we performed 3 independent MD production runs of 500 ns each (i.e. 1.5 µs accumulated time for each system). A 11 Å cutoff value was applied to Lennard-Jones and electrostatic interactions. For the PAM-ADO-A_1R-G_{i2} system, given that the MIPS521 PAM presents low binding affinity because it mostly performs weak interactions with the receptor, we applied a slight parabolic restrain with a force constant of 10 Kcal/(mol·$Å^2$) in the distance between S246(OH) and PAM(N). This avoided unbinding of the PAM ligand during the simulation time.

Metadynamics simulations

Collective Variables: Metadynamics is a powerful method to construct complex free energy landscapes of proteins as function of a few low-dimensional descriptors, also referred as collective variables (CVs).[57, 58] In this work, we selected the TM3-TM6 intracellular ends distance as the first collective variable (CV1). Specifically, we computed the center of mass (COM) distances between the backbone atoms of TM6($R105^{3.50}$, $Y106^{3.51}$, $L107^{3.52}$, $R108^{3.53}$ and $V109^{3.54}$) and TM3($Y226^{6.27}$, $G227^{6.28}$, $K228^{6.29}$, $L230^{6.30}$ and $E230^{6.31}$). For the second collective variable (CV2), we relied on the TM6 torsion, which was described by the dihedral angle formed by the alpha carbon atoms of TM7($T277^{7.42}$ and $T270^{7.35}$) and TM6 ($L236^{6.37}$, $W247^{6.48}$). These two CV were used to describe and monitor the TM6 inward-to-outward transition (A_1R activation).
Well-Tempered and multiple-walkers metadynamics simulation protocol: The PLUMED2.7[59] software package together with AMBER20[54] were used to carry out the metadynamics simulations. During a metadynamics simulation, external energy quantities (Gaussian potentials) are added at a regular number of MD steps to a selected CVs (see above). This bias potential encourages the system to escape from local energy minima and overcome energy barriers, thus allowing for an enhanced sampling of the CV conformational states.[60] After sufficient simulation time, the bias potential converges and the FEL can be reconstructed by summing the Gaussian potentials added to the CV values along the simulation time. Here, Gaussian potentials of height 0.5 kcal mol-1 and width of 0.25 (CV1) and 0.015 (CV2) were deposited every 2 ps of MD simulation at 303 K. For a smooth convergence of the bias potential we used the well-tempered (WT)[61] version of metadynamics algorithm, in which the height of the gaussian potentials were gradually decreased over time proportional to the potential deposited in the currently visited point of the CV space. A bias factor parameter of 10 was selected to control how quick the Gaussian height is decreased. In addition, the multiple-walkers approach[62] was used to improve the

conformational sampling and to speed up the metadynamics simulations. It is based on running in parallel interacting replicas (walkers) where each walker biases the identical CVs and reads the Gaussian potentials deposited by the others during the simulation, thus reconstructing the same metadynamics bias simultaneously. In particular, we run ten walker replicas. The ten walker structures (W1-10) used as starting points for the walker metadynamics simulations were carefully selected from the initial conformational sampling of the cMD simulations. Specifically, we selected 5 walker structures (W1-5) from the cMD starting from the inactive X-Ray and 5 walker structures (W6-10) from the cMD starting from the active Cryo-EM in order to provide a path of conformations that encompasses the conformational states sampled along the A_1R activation pathway **(Figure S2**). Each walker replica was run for 25 ns, giving a total of 250 ns. Finally, the FEL of the TM6 inward-to-outward transition was completely reconstructed by summing the Gaussian potentials deposited by all walker replicas as a function of the CVs.

Convergence: An indicator of convergence consists in observing that the free energy surface does not change significantly over time. We estimate the convergence of the recovered FEL monitoring the free energy difference (ΔΔG) between the local energy minima of the activation conformational surface along the simulation time. In particular, we calculated the inactive-active, the inactive-intermediate and the active-intermediate local energy minima differences (ΔΔG), see **Figure S3**. The metadynamics simulations were considered to be converged once we observed that with increasing simulation time, the energy differences between the energy minima tend to flatten. In other words, once the free energy surface does not change significantly during a relatively long period of time in the last part of the simulation. We have also assessed convergence by analyzing the CV1 values over simulation time. **Figure S4A** shows that during the first 100ns, walkers primary oscillates around their initial CV1 values. Subsequently, at around 200 ns walkers exhibit a higher frequency of crossing into regions occupied by other walkers. This is further supported by the exploration of W1 and W10, as shown in **Figure S4B**. These two walkers initially start the landscape reconstruction at the opposite extremes of the CV space. At 120 ns, they are able to escape from their respective basins and approach each other, sampling similar CV values (at approximately 240 ns). At this point of the simulation, only these two walkers have covered the entire conformational space of activation. Subsequently, they tend to return to previously sampled CV space. The observation that walkers do not become trapped in their initial CVs region, but instead explore and cross into other regions suggests that our sampling strategy, which involved starting the simulations with walkers that spanned the entire CV space of interest, has facilitated the exploration of the relevant conformational space. Although we cannot guarantee full convergence of the free energy landscape under these conditions, we successfully reconstructed the major conformational states of the receptor activation at 250 ns."

Error: We estimated the error on the 2D free energy landscape of the first collective variable (CV1), which is the TM3-TM6 intracellular ends distance (**Figure S13**) using the block averaging technique, as described in the PLUMED tutorial on calculating error bars (https://www.plumed.org/doc-v2.8/user-doc/html/lugano-4.html). We calculated the weights using the metadynamics bias potential obtained at the end of the simulation, and assuming a constant bias during the entire course of the simulation.[28] Specifically, we calculate the error using blocks of histograms of 25 ns each, covering the entire 250 ns simulation time.

Protein Energy Networks (PEN) analysis

Mean Interaction Energy Matrix (IEM): The get Residue Interaction eNergies and Networks (gRINN)[32] tool was used to calculate the pairwise residue interaction energies along the different conformational ensembles in order to obtain their respective mean interaction energy matrixes (IEM). The analysis was performed considering all residue pairs of A_1R, which resulted in 326x326 matrixes. For the complete A_1R conformational ensemble, we used one out of two conformations sampled in the metadynamics simulations resulting in 6,250 structures. For the analysis by conformational states, we split the whole conformational ensemble obtained from metadynamics simulations (12,500 structures) into the inactive, intermediate and pre-active ensembles, resulting in 3,648 structures for the inactive ensemble, 3,896 structures for the intermediate ensemble and 4,404 structures for the pre-active ensemble (**Figure S9**). The analysis of the multimeric complexes (ADO-A_1R-G_{i2}, A_1R-G_{i2} and PAM-ADO-A_1R-G_{i2}) was performed using 5,000 representative structures from the cMD. For this case, we consider both

the intra (A_1R-A_1R) and inter (A_1R-$G\alpha_{i2}$) residue pairs of interactions. In all matrixes, the positive energies (repulsive interactions) were set to 0 and the negative energies (attractive interactions) were converted to absolute values. In a second step, the matrixes were normalized. Hence, the attractive interactions were weighted containing values that range between 0 and 1.

Shortest Path Map: To study the allosteric communication by means of protein energy networks (PEN) we processed the normalized mean IEM into to Shortest Path Map (SPM) tool. [29, 33, 34] The SPM algorithm first constructs a network graph in which only the pair of residues with a normalized mean interaction energy (IE) higher than 0.1 are considered as nodes. The length of the edge connecting pair of residues (nodes) is drawn according to their normalized mean IE value (d_{ij}=-log $|IE_{ij}|$). Thus, higher normalized mean IE values (closer to 1) will have shorter edge distances, whereas lower normalized mean IE values (closer to 0.1) will have edges with long distances. At this point, we apply the Dijkstra algorithm to identify the shortest path lengths and generate the SPM graph. The Dijkstra algorithm operates through all nodes of the initial network graph and determine the shortest path to go from one node origin to all other nodes. The exploration is over when all nodes have been targeted as origin. In the SPM graph, the width of each edge and the size of each node are proportional to the number of shortest paths passing through that edge or node during the exploration. The method therefore offers the visualization of which nodes and edges are more frequently used for going through all residues of the protein, i.e. they are more central and significant for the communication pathway.

Transient pocket analysis

The MDpocket[37] tool was used to detect transient pockets in the inactive, intermediate, pre-active and fully-active ensembles. MDpocket is a fast geometry-based algorithm that relies on the concept of alpha spheres and makes extensive use of Voronoi tessellation during the cavity detection. The output provides a normalized frequency map allowing for an iso-surface representation. In the frequency map, the iso-values (Φi) ranges from 0 to 1, allowing for visualization of both permanent (Φi =1) and transient pockets (0< Φi < 1). Since we were interested in the detection of transient pockets along the receptor activation we used a Φi =0.2 iso-value for the display of pockets in all conformational ensembles studied.

AKCNOWLEDGEMENTS

This work was supported by the Bio & Medical Technology Development Program (NRF-2022M3E5F3080873), the Mid-career Researcher Program (NRF-2020R1A2C2101636), Medical Research Center (MRC) grant (NRF-2018R1A5A2025286), and the Brain Pool Program (NRF-2021H1D3A2A02038434) funded by the Ministry of Science and ICT (MSIT) through the National Research Foundation of Korea (NRF). It was also supported by the Ewha Womans University Research Grant of 2021. We are grateful to the Korea Institute of Science and Technology Information (KISTI) Supercomputing Center for providing computing resources and technical assistance (KSC-2022-CRE-0517). We thank Prof. Ferran Feixas, Dr. Raudah Lazim and Dr. Maninder for the helpful discussions and comments on the manuscript.

DATA AVAILABILITY

All the Mean Interaction energy matrixes (IEMs) computed have been deposited on GitHub (https://github.com/biochem0904/A1R-ENetDyna), which will allow users to perform customized analysis of the protein energy networks. We have also deposited the PyMOL sessions corresponding to **Figure 6** for a more comprehensive visualization and analysis.

REFERENCES

1. Wodak, S.J., et al., *Allostery in Its Many Disguises: From Theory to Applications.* Structure, 2019. **27**(4): p. 566-578.
2. Nussinov, R., C.J. Tsai, and J. Liu, *Principles of allosteric interactions in cell signaling.* J Am Chem Soc, 2014. **136**(51): p. 17692-701.
3. Tsai, C.J. and R. Nussinov, *A unified view of "how allostery works".* PLoS Comput Biol, 2014. **10**(2): p. e1003394.
4. Nussinov, R. and C.J. Tsai, *Allostery without a conformational change? Revisiting the paradigm.* Curr Opin Struct Biol, 2015. **30**: p. 17-24.
5. Guo, J. and H.X. Zhou, *Protein Allostery and Conformational Dynamics.* Chem Rev, 2016. **116**(11): p. 6503-15.
6. Dokholyan, N.V., *Controlling Allosteric Networks in Proteins.* Chem Rev, 2016. **116**(11): p. 6463-87.
7. Lu, S.Y., et al., *Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design.* Nature Communications, 2021. **12**(1): p. 4721.
8. May, L.T., et al., *Allosteric modulation of G protein-coupled receptors.* Annu Rev Pharmacol Toxicol, 2007. **47**: p. 1-51.
9. Christopoulos, A., *Allosteric binding sites on cell-surface receptors: novel targets for drug discovery.* Nat Rev Drug Discov, 2002. **1**(3): p. 198-210.
10. Erlandson, S.C., C. McMahon, and A.C. Kruse, *Structural Basis for G Protein-Coupled Receptor Signaling.* Annu Rev Biophys, 2018. **47**: p. 1-18.
11. Weis, W.I. and B.K. Kobilka, *The Molecular Basis of G Protein-Coupled Receptor Activation.* Annu Rev Biochem, 2018. **87**: p. 897-919.
12. Mafi, A., S.K. Kim, and W.A. Goddard, 3rd, *The mechanism for ligand activation of the GPCR-G protein complex.* Proc Natl Acad Sci U S A, 2022. **119**(18): p. e2110085119.
13. Bostock, M.J., A.S. Solt, and D. Nietlispach, *The role of NMR spectroscopy in mapping the conformational landscape of GPCRs.* Curr Opin Struct Biol, 2019. **57**: p. 145-156.
14. Zhou, Q., et al., *Common activation mechanism of class A GPCRs.* Elife, 2019. **8**: p. e50279.
15. Mattedi, G., et al., *A combined activation mechanism for the glucagon receptor.* Proc Natl Acad Sci U S A, 2020. **117**(27): p. 15414-15422.
16. Miao, Y. and J.A. McCammon, *Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor.* Proc Natl Acad Sci U S A, 2016. **113**(43): p. 12162-12167.

17. Fleetwood, O., J. Carlsson, and L. Delemotte, *Identification of ligand-specific G protein-coupled receptor states and prediction of downstream efficacy via data-driven modeling.* Elife, 2021. **10**: p. e60715.
18. Deb, P.K., et al., *Medicinal Chemistry and Therapeutic Potential of Agonists, Antagonists and Allosteric Modulators of A1 Adenosine Receptor: Current Status and Perspectives.* Curr Pharm Des, 2019. **25**(25): p. 2697-2715.
19. Draper-Joyce, C.J., et al., *Positive allosteric mechanisms of adenosine A1 receptor-mediated analgesia.* Nature, 2021. **597**(7877): p. 571-576.
20. Bruns, R.F. and J.H. Fergus, *Allosteric enhancement of adenosine A1 receptor binding and function by 2-amino-3-benzoylthiophenes.* Mol Pharmacol, 1990. **38**(6): p. 939-49.
21. Draper-Joyce, C.J., et al., *Structure of the adenosine-bound human adenosine A1 receptor-Gi complex.* Nature, 2018. **558**(7711): p. 559-563.
22. Glukhova, A., et al., *Structure of the Adenosine A1 Receptor Reveals the Basis for Subtype Selectivity.* Cell, 2017. **168**(5): p. 867-877 e13.
23. Nguyen, A.T.N., et al., *Role of the Second Extracellular Loop of the Adenosine A(1) Receptor on Allosteric Modulator Binding, Signaling, and Cooperativity.* Molecular Pharmacology, 2016. **90**(6): p. 715-725.
24. Do, H.N., et al., *GLOW: A Workflow Integrating Gaussian-Accelerated Molecular Dynamics and Deep Learning for Free Energy Profiling.* J Chem Theory Comput, 2022. **18**(3): p. 1423-1436.
25. Calvo-Tusell, C., et al., *Time Evolution of the Millisecond Allosteric Activation of Imidazole Glycerol Phosphate Synthase.* J Am Chem Soc, 2022. **144**(16): p. 7146-7159.
26. Maria-Solano, M.A., J. Iglesias-Fernandez, and S. Osuna, *Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution.* J Am Chem Soc, 2019. **141**(33): p. 13049-13056.
27. Kuzmanic, A., et al., *Changes in the free-energy landscape of p38alpha MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations.* Elife, 2017. **6**: p. e22175.
28. Branduardi, D., G. Bussi, and M. Parrinello, *Metadynamics with Adaptive Gaussians.* J Chem Theory Comput, 2012. **8**(7): p. 2247-54.
29. Maria-Solano, M.A., et al., *In Silico Identification and Experimental Validation of Distal Activity-Enhancing Mutations in Tryptophan Synthase.* ACS Catal, 2021. **11**(21): p. 13733-13743.
30. Kong, Y. and M. Karplus, *The signaling pathway of rhodopsin.* Structure, 2007. **15**(5): p. 611-23.
31. Lee, Y., S. Choi, and C. Hyeon, *Mapping the intramolecular signal transduction of G-protein coupled receptors.* Proteins, 2014. **82**(5): p. 727-43.
32. Sercinoglu, O. and P. Ozbek, *gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations.* Nucleic Acids Res, 2018. **46**(W1): p. W554-W562.
33. Romero-Rivera, A., M. Garcia-Borras, and S. Osuna, *Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity.* ACS Catal, 2017. **7**(12): p. 8524-8532.
34. Osuna, S., *The challenge of predicting distal active site mutations in computational enzyme design.* Wiley Interdisciplinary Reviews-Computational Molecular Science, 2021. **11**(3): p. e1502.
35. Miao, Y., et al., *Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor.* Sci Rep, 2018. **8**(1): p. 16836.
36. Nguyen, A.T.N., et al., *Extracellular Loop 2 of the Adenosine A(1) Receptor Has a Key Role in Orthosteric Ligand Affinity and Agonist Efficacy.* Molecular Pharmacology, 2016. **90**(6): p. 703-714.
37. Schmidtke, P., et al., *MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories.* Bioinformatics, 2011. **27**(23): p. 3276-3285.
38. Kruse, A.C., et al., *Activation and allosteric modulation of a muscarinic acetylcholine receptor.* Nature, 2013. **504**(7478): p. 101-6.
39. Cheng, R.K.Y., et al., *Structural insight into allosteric modulation of protease-activated receptor 2.* Nature, 2017. **545**(7652): p. 112-115.
40. Tan, Q.X., et al., *Structure of the CCR5 Chemokine Receptor-HIV Entry Inhibitor Maraviroc Complex.* Science, 2013. **341**(6152): p. 1387-1390.

41. Christopher, J.A., et al., *Fragment and Structure-Based Drug Discovery for a Class C GPCR: Discovery of the mGlu5 Negative Allosteric Modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile).* J Med Chem, 2015. **58**(16): p. 6653-64.
42. Christopher, J.A., et al., *Structure-Based Optimization Strategies for G Protein-Coupled Receptor (GPCR) Allosteric Modulators: A Case Study from Analyses of New Metabotropic Glutamate Receptor 5 (mGlu5) X-ray Structures.* J Med Chem, 2019. **62**(1): p. 207-222.
43. Fenalti, G., et al., *Molecular control of delta-opioid receptor signalling.* Nature, 2014. **506**(7487): p. 191-6.
44. Lu, J., et al., *Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40.* Nat Struct Mol Biol, 2017. **24**(7): p. 570-577.
45. Oswald, C., et al., *Intracellular allosteric antagonism of the CCR9 receptor.* Nature, 2016. **540**(7633): p. 462-465.
46. Srivastava, A., et al., *High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875.* Nature, 2014. **513**(7516): p. 124-7.
47. Zhang, D., et al., *Two disparate ligand-binding sites in the human P2Y1 receptor.* Nature, 2015. **520**(7547): p. 317-21.
48. Li, Y., et al., *The full activation mechanism of the adenosine A(1) receptor revealed by GaMD and Su-GaMD simulations.* Proc Natl Acad Sci U S A, 2022. **119**(42): p. e2203702119.
49. Rasmussen, S.G., et al., *Crystal structure of the beta2 adrenergic receptor-Gs protein complex.* Nature, 2011. **477**(7366): p. 549-55.
50. Carpenter, B., et al., *Erratum: Structure of the adenosine A2A receptor bound to an engineered G protein.* Nature, 2016. **538**(7626): p. 104-107.
51. Cheng, R.K.Y., et al., *Structures of Human A1 and A2A Adenosine Receptors with Xanthines Reveal Determinants of Selectivity.* Structure, 2017. **25**(8): p. 1275-1285 e4.
52. Soundararajan, M., et al., *Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(17): p. 6457-6462.
53. Eswar, N., et al., *Comparative protein structure modeling using Modeller.* Curr Protoc Bioinformatics, 2006. **Chapter 5**: p. Unit-5 6.
54. D.A. Case, K.B., I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz,R Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman, *Amber 2020.* University of California, San Francisco, 2020.
55. M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. H., *Gaussian 09, Revision A.02. Gaussian, Inc., Wallingford CT 2016*
56. Mayol, E., et al., *HomolWat: a web server tool to incorporate 'homologous' water molecules into GPCR structures.* Nucleic Acids Res, 2020. **48**(W1): p. W54-W59.
57. Bussi, G. and A. Laio, *Using metadynamics to explore complex free-energy landscapes.* Nature Reviews Physics, 2020. **2**(4): p. 200-212.
58. Laio, A. and F.L. Gervasio, *Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science.* Reports on Progress in Physics, 2008. **71**(12): p. 126601.
59. Tribello, G.A., et al., *PLUMED 2: New feathers for an old bird.* Computer Physics Communications, 2014. **185**(2): p. 604-613.
60. Laio, A. and M. Parrinello, *Escaping free-energy minima.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(20): p. 12562-12566.
61. Barducci, A., G. Bussi, and M. Parrinello, *Well-tempered metadynamics: A smoothly converging and tunable free-energy method.* Physical Review Letters, 2008. **100**(2): p. 020603.
62. Raiteri, P., et al., *Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics.* Journal of Physical Chemistry B, 2006. **110**(8): p. 3533-3539.

FIGURES

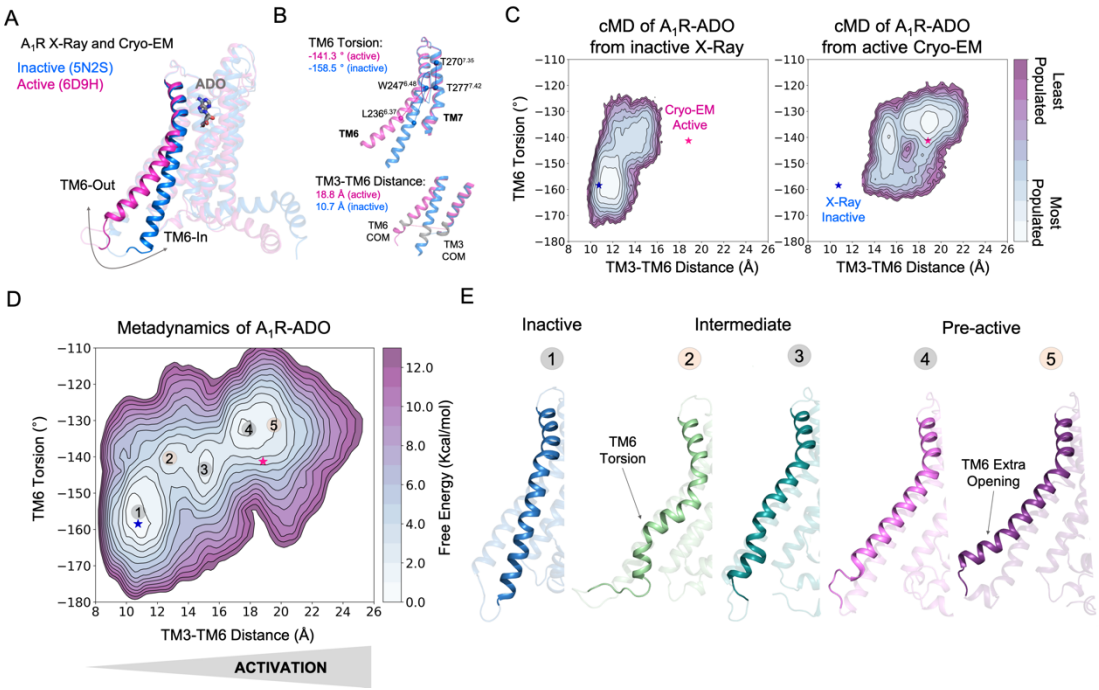

Figure 1. Free energy landscape (FEL) of A_1R activation in presence of adenosine (ADO). (A) TM6 inward-to-outward conformational transition observed in the inactive (PDB 5N2S) and active (PDB 6D9H) X-Ray and Cryo-EM structures. **(B)** Features used to follow the TM6 inward-to-outward transition (receptor activation) in this work. The TM6 torsion corresponds to the dihedral angle formed by the alpha carbon atoms of $L236^{6.37}$, $W247^{6.48}$, $T277^{7.42}$ and $T270^{7.35}$. For the TM3-TM6 intracellular ends distance, we computed the center of mass (COM) distances between the backbone atoms of TM3($Y226^{6.27}$, $G227^{6.28}$, $K228^{6.29}$, $L230^{6.30}$ and $E230^{6.31}$) and TM6($R105^{3.50}$, $Y106^{3.51}$, $L107^{3.52}$, $R108^{3.53}$ and $V109^{3.54}$). The X-ray and Cryo-EM values are also shown. **(C)** Population analysis obtained from conventional molecular dynamics (cMD) simulations of A_1R activation starting from the inactive X-Ray and active Cryo-EM structures, the inactive X-ray and active Cryo-EM coordinates are projected as blue and magenta stars, respectively. **(D)** Reconstruction of the FEL associated with the A_1R activation obtained from metadynamics simulations. The most relevant conformational states are labeled from 1 to 5. Note that the lowest energy states (1,3 and 4) are labeled in gray while the other (2 and 5) in orange. The X-ray and Cryo-EM coordinates are also projected. **(E)** Representative structures of the inactive (1), Intermediate (2-3) and Pre-active (4-5) conformational states sampled along A_1R-ADO activation.

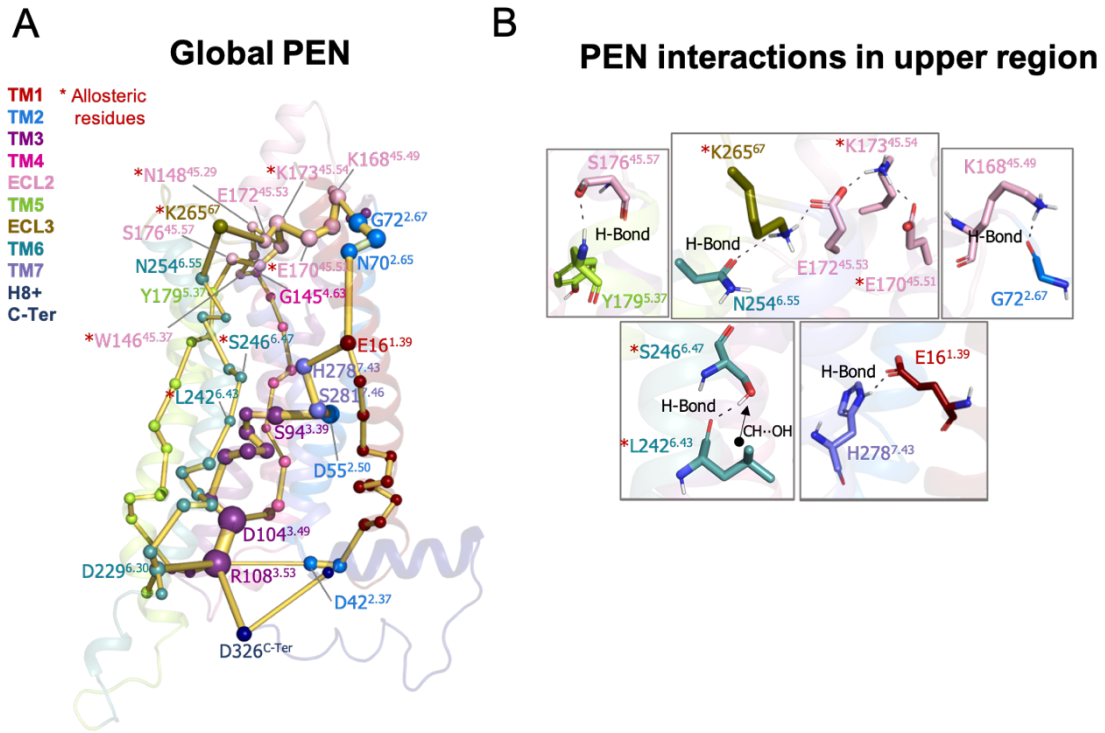

Figure 2. Protein energy networks (PEN) of A_1R-ADO conformational ensemble. The PEN identifies extra and intracellular communication centers together with the allosteric pathways that interconnect them. The PEN residues (nodes) are represented by colored spheres as a function of the receptor region (e.g. TM6 nodes in teal) while the allosteric pathways (edges) as yellow-orange sticks. The size of each edge and node correspond to their importance for the allosteric communication. The experimentally identified allosteric residues captured in the PEN nodes are labeled with a red asterisk. **(B)** Relevant interactions of the PEN in the upper region of the receptor.

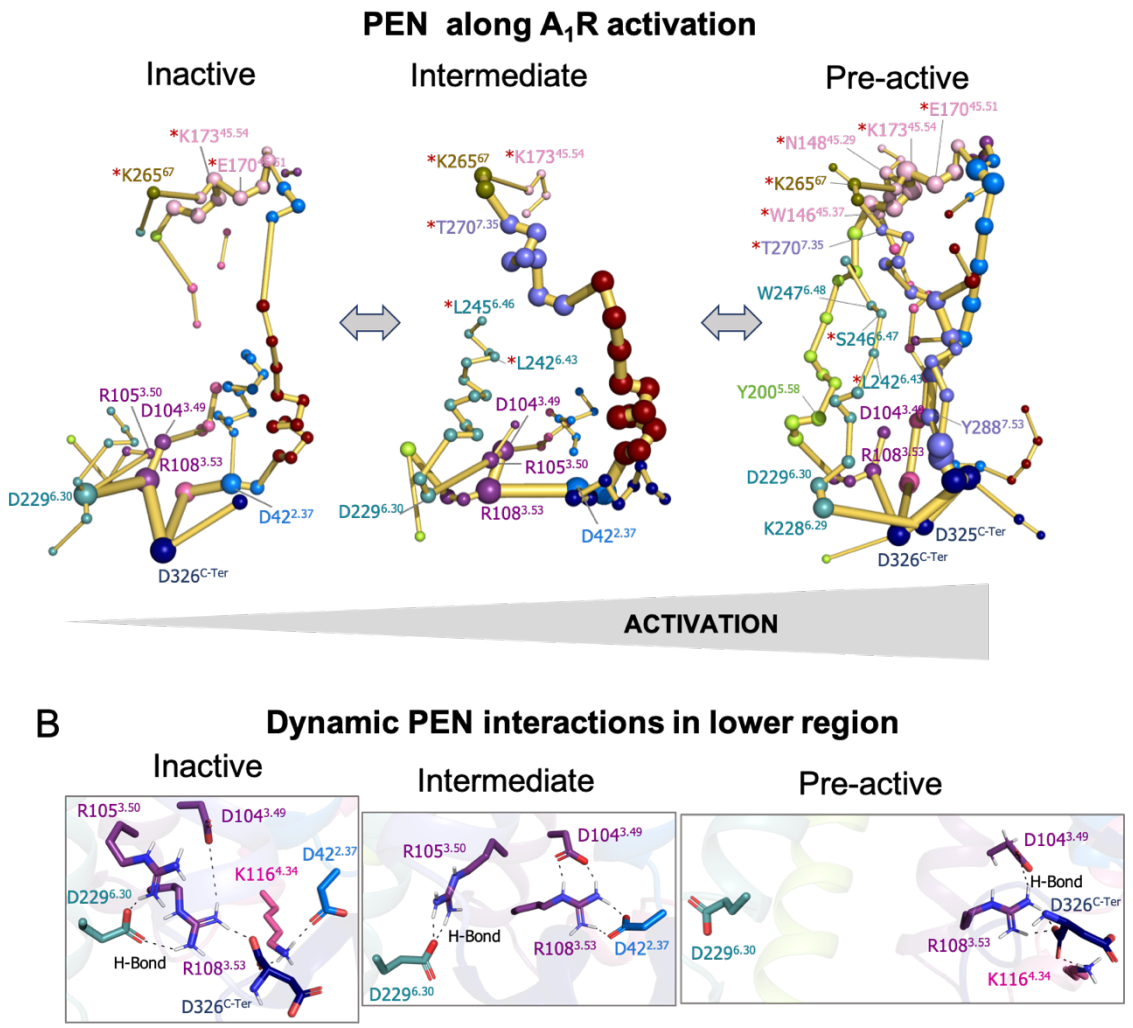

Figure 3. Protein energy networks (PEN) of A_1R-ADO in the inactive, intermediate and pre-active states. (A) The PEN residues (nodes) are represented by colored spheres as a function of the receptor region (e.g. TM6 nodes in teal) while the allosteric pathways (edges) as yellow-orange sticks. The size of each edge and node correspond to their importance for the allosteric communication. The experimentally identified allosteric residues captured in the PEN nodes are labeled with a red asterisk. The allosteric communication is enhanced along the receptor activation. **(B)** Relevant interactions found in the PEN of the lower region of the receptor that are altered along activation.

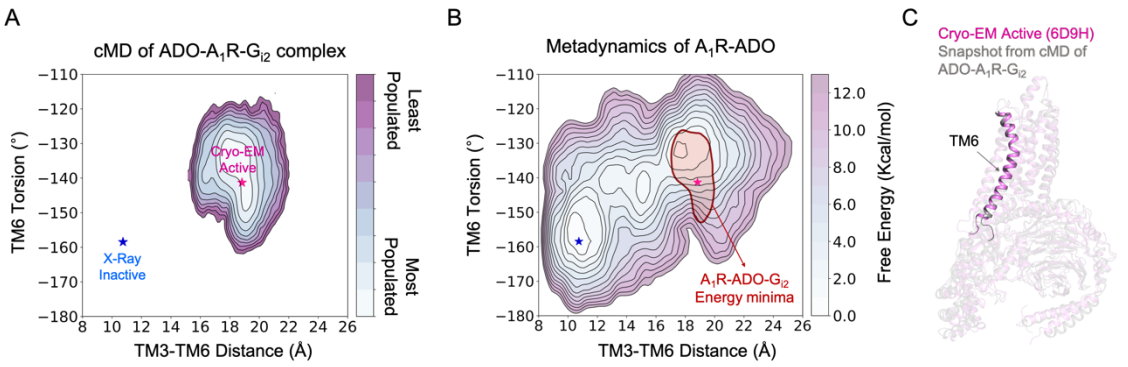

Figure 4. Effect of G-proteins binding in the A_1R-ADO conformational ensemble. (A) Population analysis of A_1R activation in the ADO-A_1R-G_{i2} complex obtained from conventional molecular dynamics (cMD) simulations. The TM6 torsion corresponds to the dihedral angle formed by the alpha carbon atoms of L236$^{6.37}$, W247$^{6.48}$, T277$^{7.42}$ and T270$^{7.35}$. For the TM3-TM6 intracellular ends distance, we computed the center of mass (COM) distances between the backbone atoms of TM3(Y226$^{6.27}$, G227$^{6.28}$, K228$^{6.29}$, L230$^{6.30}$ and E230$^{6.31}$) and TM6(R105$^{3.50}$, Y106$^{3.51}$, L107$^{3.52}$, R108$^{3.53}$ and V109$^{3.54}$). The inactive and active X-ray and Cryo-EM coordinates are projected as blue and magenta stars, respectively. **(B)** Projection of the ADO-A_1R-G_{i2} energy minima obtained from cMD over the FEL associated with the A_1R activation obtained from metadynamics simulations. The ADO-A_1R-G_{i2} energy minima (depicted in red) is centered on the coordinates of the active Cryo-EM structure. **(C)** Overlay of the active Cryo-EM structure (PDB 6D9H) and a representative snapshot from the ADO-A_1R-G_{i2} energy minima.

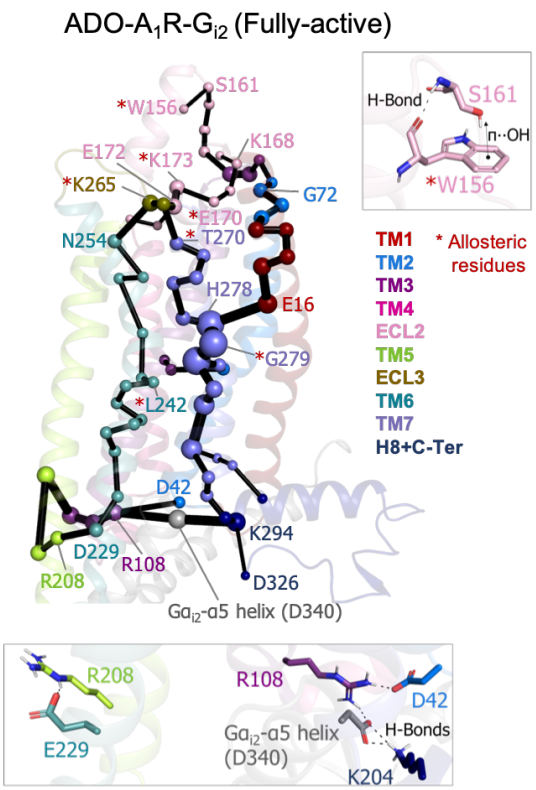

Figure 5. Effect of G-proteins binding in A_1R-ADO Protein Energy Networks (PEN). The PEN residues (nodes) are represented by colored spheres as a function of the receptor region (e.g. TM6 nodes in teal) while the allosteric pathways (edges) as yellow-orange sticks. The size of each edge and node correspond to their importance for the allosteric communication. The experimentally identified allosteric residues captured in the PEN nodes are labeled with a red asterisk. Relevant interactions of the PEN in the upper and lower regions of the receptor are also shown.

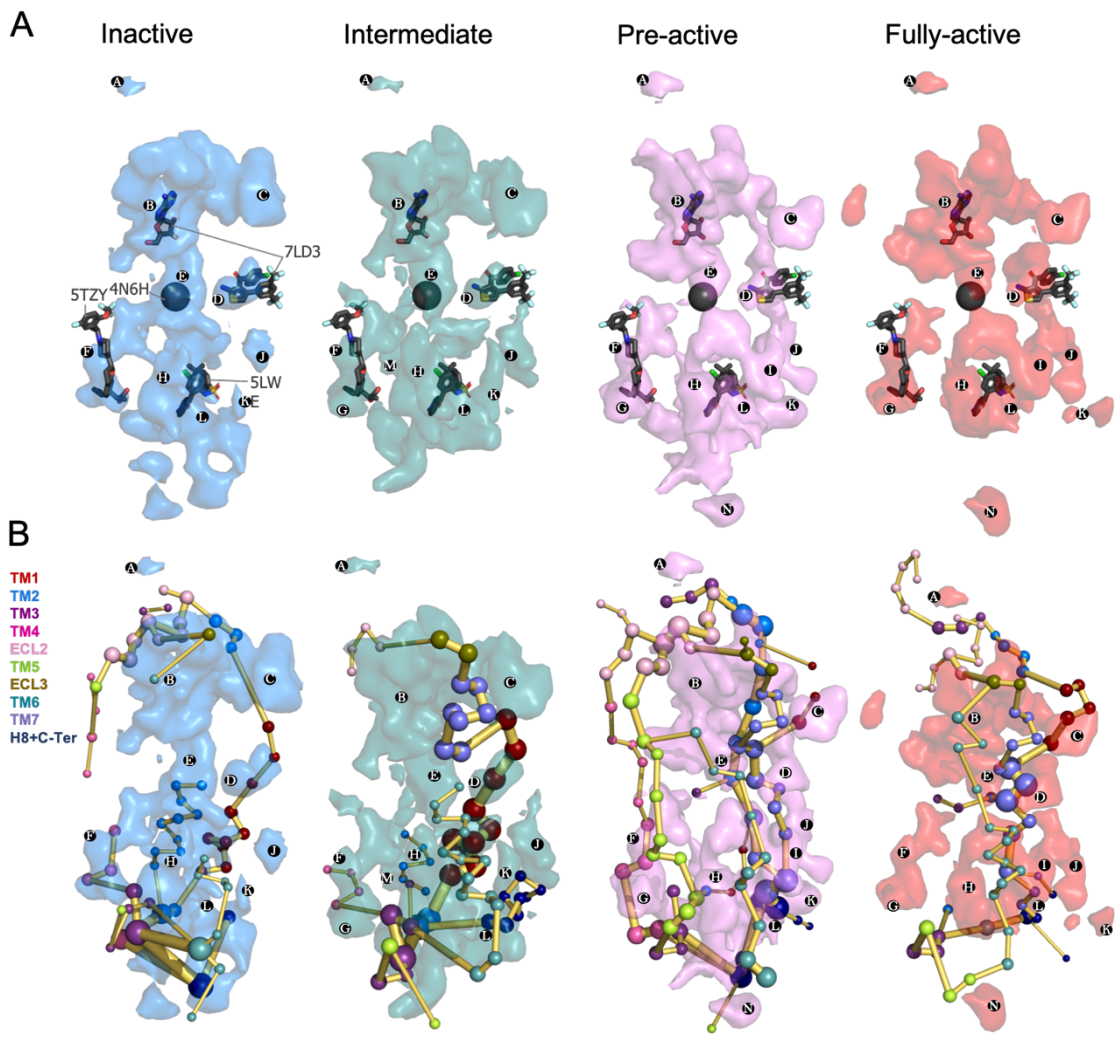

Figure 6. Energy coupling between the transient pockets formed along activation. (A) Iso-surface representation of the normalized frequency map (set at Φi =0.2 iso-value) obtained from MDPocked in the inactive (blue), intermediate (teal), pre-active (violet) and fully-active (red) ensembles. The transient pockets are labeled from A to N. The allosteric modulators that overlap with the pockets found are shown as dark gray sticks (PDB 7DL3, 5TZY and 5LWE) and spheres (PDB 4N6H). **(B)** Overlap of the transient pockets and the protein energy networks (PEN). The PEN residues (nodes) are represented by colored spheres as a function of the receptor region (e.g. TM6 nodes in teal) while the allosteric pathways (edges) as yellow-orange sticks.

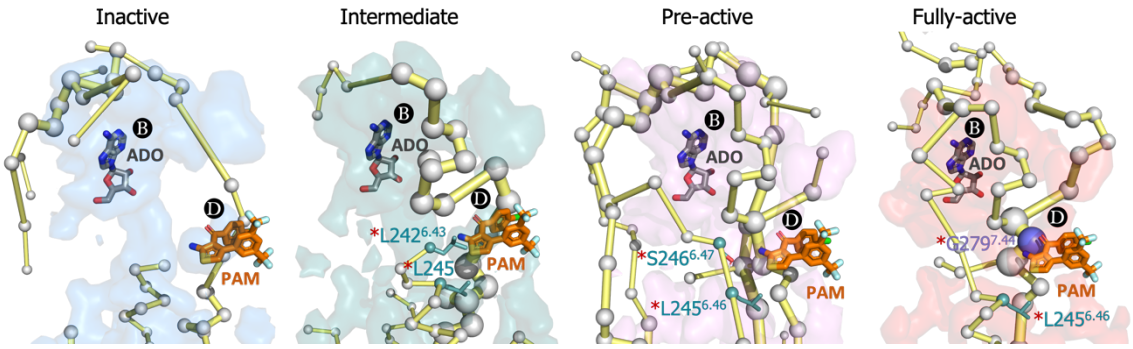

Figure 7. Energy coupling between PB and PD along activation. Zoom view of the transient pockets and protein energy networks of the upper region of the receptor. Adenosine (ADO) in pocket B and the MIPS521 positive allosteric modulator (PAM) in pocket D, both aligned from PDB 7LD3, are depicted in gray and orange sticks, respectively. The PEN residues (nodes) are represented by light gray spheres. The experimentally identified allosteric residues located in pocked D that affects the PAM are colored as a function of the receptor region (TM6 nodes in teal and TM7 nodes in purple) and highlighted with a red asterisk. The allosteric pathways (edges) are depicted as yellow-orange sticks.

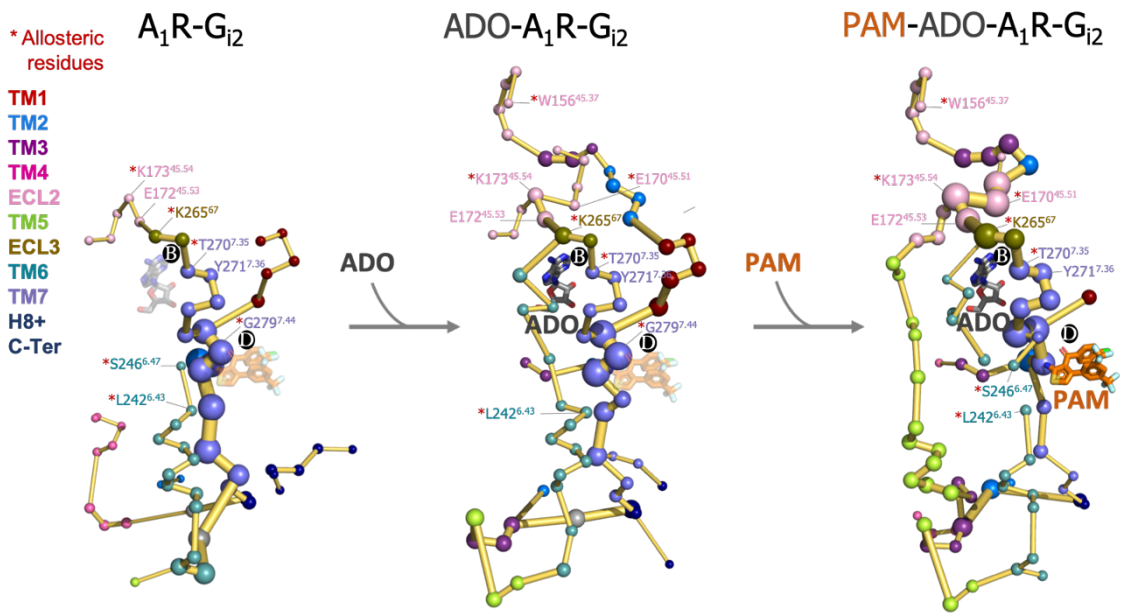

Figure 8. Effect of ADO and MIPS51 PAM binding in the A_1R-G_{i2} Protein Energy Networks (PEN). Adenosine (ADO) in pocket B and the MIPS521 positive allosteric modulator (PAM) in pocket D, both aligned from PDB 7LD3, are depicted in gray and orange sticks, respectively. Note that ADO and PAM sticks are displayed with transparency in the systems in which they are absent. The PEN residues (nodes) are represented by colored spheres as a function of the receptor region (e.g. TM6 nodes in teal) while the allosteric pathways (edges) as yellow-orange sticks. The experimentally identified allosteric residues captured in the PEN nodes are labeled with a red asterisk.