PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Mapping the intramolecular signal transduction of G-protein coupled receptors

Yoonji Lee,[1] Sun Choi,[1]* and Changbong Hyeon[2]*

[1] National Leading Research Lab (NLRL) of Molecular Modeling and Drug Design, College of Pharmacy, Graduate School of Pharmaceutical Sciences, and Global Top 5 Research Program, Ewha Womans University, Seoul 120-750, Korea

[2] School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Korea

ABSTRACT

G-protein coupled receptors (GPCRs), a major gatekeeper of extracellular signals on plasma membrane, are unarguably one of the most important therapeutic targets. Given the recent discoveries of allosteric modulations, an allosteric wiring diagram of intramolecular signal transductions would be of great use to glean the mechanism of receptor regulation. Here, by evaluating betweenness centrality (C_B) of each residue, we calculate maps of information flow in GPCRs and identify key residues for signal transductions and their pathways. Compared with preexisting approaches, the allosteric hotspots that our C_B-based analysis detects for A_{2A} adenosine receptor (A_{2A}AR) and bovine rhodopsin are better correlated with biochemical data. In particular, our analysis outperforms other methods in locating the rotameric microswitches, which are generally deemed critical for mediating orthosteric signaling in class A GPCRs. For A_{2A}AR, the inter-residue cross-correlation map, calculated using equilibrium structural ensemble from molecular dynamics simulations, reveals that strong signals of long-range transmembrane communications exist only in the agonist-bound state. A seemingly subtle variation in structure, found in different GPCR subtypes or imparted by agonist bindings or a point mutation at an allosteric site, can lead to a drastic difference in the map of signaling pathways and protein activity. The signaling map of GPCRs provides valuable insights into allosteric modulations as well as reliable identifications of orthosteric signaling pathways.

Proteins 2013; 00:000–000.

Key words: GPCR; adenosine receptors; allostery; network theory; betweenness centrality.

INTRODUCTION

G-protein coupled receptors (GPCRs), a major gatekeeper on the cell surface, mediate various physiological processes, such as vision, olfaction, cardiovascular function, and immune responses, which makes GPCRs one of the most important therapeutic targets[1]. Consisting of seven α-helical transmembrane (TM) domains, extracellular and intracellular loops (ECLs and ICLs), GPCRs relay extracellular signals to the cytoplasmic domain and activate proteins associated with signal transduction pathways.[1,2] The activity of GPCRs is highly selective to the type of extracellular signals,[3] and is sensitively modulated by point mutations,[4,5] the latter of which is closely related to the development of drug resistance as well.[6,7] Because agonist binding to orthosteric sites enables accommodation and activation of G-protein by regulating the conformational change in cytoplasmic domain,[1] developing antagonist or agonist drugs targeting at orthosteric sites has been a straightforward strategy of drug design. Such strategy, however, has often shown limited success due to the high sequence conservation among the members of a GPCR subfamily. Instead, there have been several reports on the efficacy of allosteric modulators in ion-channels[8] and other systems[9] as well as GPCRs,[10] which highlights the role of allosteric sites in regulating the orthosteric signaling.

Both orthosteric signaling and allosteric modulation are associated with long-range communications between two remote sites in the receptor structure. Such communications, which are altogether referred to as allostery,

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Leading Research Lab (NLRL) program funded by the Ministry of Science, ICT & Future Planning (MSIP) and the National Research Foundation of Korea (NRF); Grant number: 2011-0028885.

*Correspondence to: Sun Choi, Ewha Womans University, Seoul 120-750, Korea, E-mail: sunchoi@ewha.ac.kr or Changbong Hyeon, School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Korea, E-mail: hyeoncb@kias.re.kr

Received 8 August 2013; Revised 13 September 2013; Accepted 10 October 2013
Published online 25 October 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24451

could be a consequence of a special balance of intramolecular forces formed in the network of the inter-residue contacts. Even though overall backbone topologies are similar between two different receptors belonging to the same subfamily, the sequence variation alters the connectivity map or local packing, which could be led to drastic differences in the allosteric signaling map. Despite long history of study on protein dynamics and even with the atomistic details of three-dimensional structure at hand, the structural basis of protein allostery still remains elusive and stubbornly resists revealing its microscopic underpinnings. While protein mutagenesis is a standard experimental method to identify key residues for protein function, the associated experiments are laborious and time-consuming. To overcome this experimental difficulty, there has been a growing interest in the use of theoretical/bioinformatics analysis. Careful statistical evaluations of multiple sequence alignment (MSA) of a protein family can be used to detect a set of evolutionally correlated residues.[11–14] In addition, variants of normal mode analysis have been proposed to identify key residues that control "functional" motion of enzymes and molecular motors.[15–18] Although a complementary use of these methods with molecular simulations would hold good promise to decipher the allosteric network of residues that are critical for the functional dynamics of proteins,[19] a certain class of residues are still difficult to identify if the sequence conservation of the residues is too strong, or if the residues, deeply buried at the core regions of proteins, show only a minor conformational change along with the global conformational dynamics of protein. For instance, in class A GPCRs, 18 key residues buried in the TM region, called microswitches (or rotamer toggle switches)[1,20] belong to such a class (see below).

Here, we propose a simple but powerful method to calculate the map of allosteric signal flow within protein structure, which identifies core allosteric sites including the above-mentioned microswitches. For a protein structure represented as a network of residues, we used a measure in the network theory called "betweenness centrality" (C_B)[21,22] to evaluate the importance of each residue from the perspective of the flow of information (see Materials and Methods). By adopting A_{2A} adenosine receptor (A_{2A}AR) and other GPCRs as model systems, we decide residues important for the allosteric signaling and pathways of signal flow. The comparison of the results from our analysis with those from other methods shows that the C_B-based network analysis of protein structure is much simpler, but is more reliable in identifying the allosteric hotspots that includes microswitches. Furthermore, allosteric hotspots are identified from another analysis adapting the concept of network vulnerability;[23,24] and explicit calculation of individual multiple pathways linking the clusters of long-range correlated residues across transmembrane shows that majority of paths pass through the hotspot residues we predicted. The predictions from the C_B-based network analysis of protein structure should be of great use not only to complement mutagenesis study but also to elucidate the origin of subtype selectivity as well as the activation and regulation mechanisms of GPCRs.

MATERIALS AND METHODS

Quantification of sequence conservation

For a given multiple sequence alignment (MSA) of a protein family, the following statistical free energy-like function scaled by an arbitrary energy scale k_BT^* quantifies the extent of sequence conservation:[13,19]

$$\Delta G_i/k_BT^* = \sqrt{\frac{1}{C_i}\sum_{\alpha=1}^{20}\left[p_i^\alpha \log\frac{p_i^\alpha}{p_\alpha}\right]^2}, \qquad (1)$$

where C_i is the number of amino acid types at position i along the sequence, α denotes amino acid species, p_i^α is the frequency of an amino acid α at the position i, and p_α is the frequency of an amino acid α in the full MSA, which serves as the background frequency. Note that the quantity "$S=\sum_{\alpha=1}^{20} p_i^\alpha \log(p_i^\alpha/p_\alpha)$" is the relative entropy; $S=0$ if p_i^α is no different than p_α for all α. The larger the value of ΔG_i, the sequence at the position i is better conserved. In this article, we computed $\Delta G(\mathrm{GPCR})/k_BT^*$ and $\Delta G(\mathrm{AR})/k_BT^*$, each of which is evaluated using different MSA. To obtain the MSAs of AR and GPCR families, we collected the sequences of adenosine receptor family (219 sequences) and class A GPCR family (26,655 sequences) from UniProtKB and Pfam database, respectively. After filtering the redundancy, we obtained 208 sequences and 24,507 sequences for AR and GPCR family, respectively. For GPCR, sequence clustering was performed with 40% identity to reduce the sequence space size, and 2,471 sequences were obtained. Based on these sequences, the multiple sequence alignment (MSA) was produced using the log-expectation (MUSCLE) program.[25]

Generating the minimal energy structures and conformational ensemble of the human A_{2A} adenosine receptor

The conformational flexibility of GPCRs makes it difficult to obtain high-resolution X-ray crystal structures, particularly, in the active state. Although several X-ray crystal structures of the A_{2A}AR are determined in their antagonist or agonist-bound forms,[26,27] structural information in the apo form or fully active state is not yet available.[28] To prepare the human A_{2A}AR models (residues from I3 to Q310) including all the loop regions, homology modeling was performed using MODELER

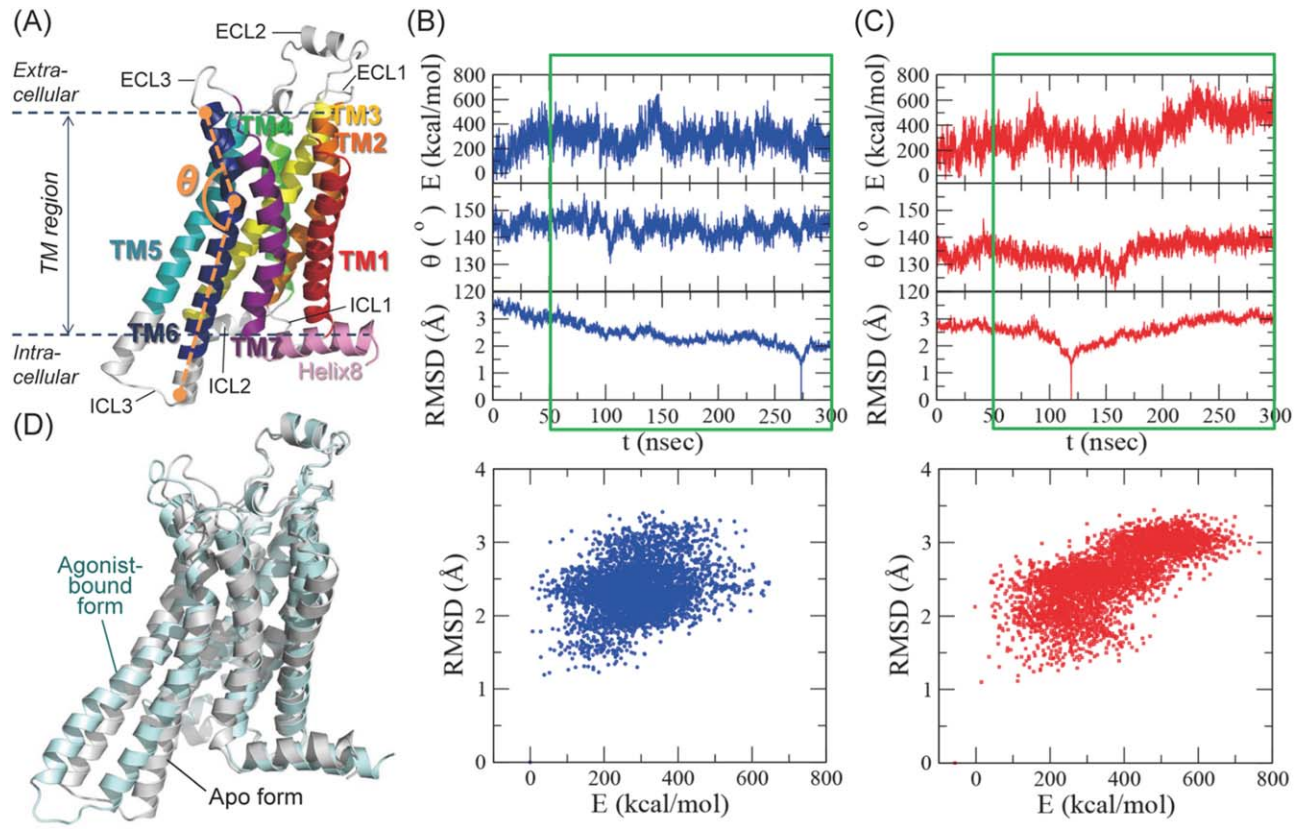


Figure 1

Structure and dynamics of A_{2A} adenosine receptor. (**A**) Seven TM helices and the intra- and extracellular loops. (**B**, **C**) Total conformational energy, TM6 tilt angle (θ) measured between three points defined along the center of helix using three group of residues (255–258, 244–247, 219–222), and RMSD in reference to the minimum energy structure from the MD trajectories of the (B) apo and (C) agonist-bound forms. Analysis was carried out for the boxed time interval, which excludes the first 50 nsec trajectories. (**D**) The minimum energy structures of the A_{2A}ARs in the apo and agonist-bound forms are overlaid to show that the most significant difference between the two forms is in the cytoplasmic region of TM5-ICL3-TM6. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

program implemented in Discovery Studio v.3.1 (Accelrys Inc., San Diego, CA, USA). We used the structures with PDB IDs, 3EML,[27] and 3QAK[26] as templates for the apo and agonist-bound forms, respectively, and 2YDV[29] and 3PWH to generate models for the loop regions that were not determined in 3EML and 3QAK. Conserved disulfide bridges, C71-C159, C74-C146, C77-C166, and C259-C262, were retained, and the agonist ligand was inserted to the agonist-bound-form model. The models were optimized with simulated annealing and selected based on the DOPE score. The final homology structures were obtained under GBSW implicit solvent hamiltonian by using conjugate gradient method. To generate the minimum energy structures and thermal structural ensemble of the A_{2A}ARs, we performed molecular dynamics simulation for 300 nsec with the NAMD v2.8 package using the CHARMM22/CMAP force field.[30] To construct an explicit membrane system, the TM region of the A_{2A}AR was predicted based on the Orientations of Proteins in Membranes (OPM) database and the palmitoyloleoylphosphatidylcholine (POPC) membrane was placed around the TM region of the receptor. Then, the receptor in membrane system was solvated with the explicit water molecules and ionized with 150 mM KCl. The whole system was energy minimized in the order of lipid membrane, waters, and the entire molecules, followed by the heating, equilibration and production runs for 300 nsec under NPT ensemble. The trajectories of production run were monitored in terms of total conformational energy, tilt angle of TM6, and root mean square deviation relative to the initial ($t=0$) structure. In accord with the common notion for GPCR dynamics, the tilt angle of TM6 varied between 135° and 150° for the apo form, and between 120° and 145° for agonist-bound form [Figs. 1(B,C)]. Finally, the minimal energy conformations from the simulated trajectories were obtained for the apo and agonist-bound forms. Our minimal energy conformation for the agonist-bound form has a tilt angle 133° in TM6, whereas the agonist-bound crystal structure 3QAK has a tilt angle of 142°

Statistical assessment of a prediction method

Since there are 18 microswitches in class A GPCRs, the probability (p_m) of correctly identifying at least one microswitch out of 308 residues of GPCRs is given by $p_m = \frac{18}{308} \approx 0.06$. Then the expectation value of identifying microswitches by randomly selecting n residues is $\langle n \rangle_{rand} = n \times p_m$. Thus, if N_m microswitches are identified with a certain method, one can evaluate prediction efficiency of the method by calculating the ratio between N_m and $\langle n \rangle_{rand}$, i.e.,

$$\varphi_m = \frac{N_m}{\langle n \rangle_{rand}}. \quad (2)$$

Construction of the residue interaction network

We constructed the residue interaction network by representing each amino acid residue as a single node. To take into account the effect of side chain, we considered two coarse-grained centers per residue, i.e., Cα carbon for backbone and a farthest heavy atom from Cα for the side chain. By doing so, we included the cases of backbone-backbone, backbone-side chain, side chain-side chain contacts. In our network model, a link was established between two Cα-carbons when any pair of backbone and side chain of two residues is less than 7 Å,[31] thus the side chains are implicit in the network.

Network centralities

Simplifying architecture of complex system into a network (graph), which is represented with "nodes" (vertices) and "links" (edges), can be used as a powerful tool to extract key properties of the system topology and its components.[32] Originally devised for analyzing social phenomena and later actively extended to reveal hub proteins central to the cellular, regulatory, metabolic networks as well as network property of each organism,[33–35] network analysis can be carried out for studying protein structures as well. In the last decade, much attention has been paid in this direction. As a general statistical property of protein structure networks, networks of folded proteins display small-worldness, but are not scale-free.[36–38] By quantifying key network properties for monomeric protein structures, one can address issues such as the plasticity of protein structures, folding of protein domains, and identify key residues along the folding pathways.[39–42] In fact, the network analysis of protein structures can be extended further to identify key residues for allostery and their wiring diagram. Several studies have recently been carried out to address the microscopic mechanism of protein allostery by applying the strategies of network or community analysis in conjunction with molecular dynamics simulation on model systems[43,44] including GPCRs.[45] To address the issue of allostery, we utilized the betweenness centrality, one of the most fundamental concepts in network analysis explained below, in identifying allosteric hotspots by surmizing that allosteric hotspots are the mediators of information flow in a network topology of a given protein structure.

Here the definitions are given for the three representative types of centrality for a node in a network: (1) The degree centrality $C_D(v)$ measures the number of edges linked to a node v.

$$C_D(v) = deg(v). \quad (3)$$

Note that $C_D(v)$ is identical to the number of contacts with its neighboring residues. (2) The closeness centrality $C_C(v)$, an inverse of mean geodesic distance (shortest path length) from all other nodes to the node v, measures how fast a signal from the node v can be transmitted to other nodes.

$$C_C(v) = \left(\sum_{i=1}^{N} d(i,v)/(N-1) \right)^{-1}, \quad (4)$$

where $d(i,v)$ is the minimal number of edges that bridge the nodes i and v. For a given network topology, $d(i,j)$ can be calculated by using Dijkstra's algorithm.[46] (3) The betweenness centrality is the measure of the extent to which a node has control over transmission of information between the nodes in the network, which is defined as:[22]

$$C_B(v) = \frac{2}{(N-1)(N-2)} \sum_{s=1}^{N-1} \sum_{t=s+1}^{N} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad (5)$$

where $s \neq t \neq v$. In the above definition, σ_{st} is the number of shortest paths linking the nodes s and t, and $\sigma_{st}(v)$ is the number of shortest paths linking the nodes s and t via the node v.[22] The factor $\frac{(N-1)(N-2)}{2}$ is the normalization constant. To calculate $C_B(v)$, we used Brandes algorithm,[47] which can reduce the computational cost of Eq. (5) substantially. The significance of betweenness centrality is succinctly illustrated in Figure 2 using a graph where both C_D and C_B values are computed at each node. The node x has a greater connectivity (C_D=6) to other nodes but its removal from the network does not destroy the communication among other nodes. In contrast the node y has less connectivity (C_D=4) than x; yet upon removal of y the whole graph would be split into three pieces. In the light of communication or the flow of information the node y is most critical. Note that y has the highest C_B value among the whole nodes. Although a few studies[24,48–50] might appear similar in spirit to our work in that they also use

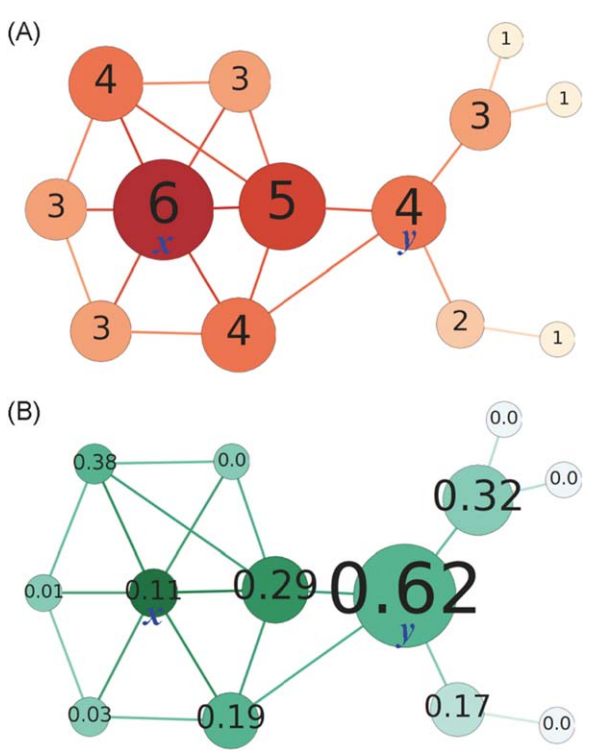


Figure 2

An example of graph showing the difference between the degree and betweenness centralities. (**A**) Degree centrality, (**B**) betweenness centrality. The calculated centrality value is marked in each node. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

centrality measures and shortest paths to decipher the allostery, it should be noted that different centrality measure has different assessment of each node. The betweenness centrality, which evaluates the importance of each node based on the amount of traffic or the amount of inter-node communication, is one of the most ideal measures to identify allosteric hotspots for a given protein structure.

RESULTS AND DISCUSSION

Microswitches: Benchmarks for prediction tools on GPCR allostery

The activation mechanism of receptors belonging to class A GPCRs, which include adenosine, $\beta_{1,2}$-adrenergic, rhodopsin, chemokine, dopamine, histamine receptors, is believed to be accompanied by a global rearrangement of TM helices that helps accommodate the binding of G-proteins. In particular, the newly resolved X-ray crystal structure of the active form of β_2-adrenergic receptor complexed with heterotrimeric G-protein[51] has lent strong support on such proposal by clearly demonstrating that 10^o outward tilt of the intracellular part of TM6 helix is essential for the full activation of the receptor. For the class A GPCRs, it has been suggested that the activation mechanism is regulated by 18 microswitches (N24, D52, D101, R102, Y103, W129, P189, Y197, E228, C245, W246, P248, N280, S281, N284, P285, Y288, F295),[20,52] which consist of DRY ($D101^{3.49}$, $R102^{3.50}$, and $Y103^{3.51}$ in TM3), CWxP ($C245^{6.47}$, $W246^{6.48}$, and $P248^{6.50}$ in TM6), and NPxxY ($N284^{7.49}$, $P285^{7.50}$, and $Y288^{7.53}$ in TM7) motifs[53,54] (where "x" stands for any amino acid residue and the numbers in the superscript of residues are based on the Ballesteros Weinstein numbering system[55]), and others. Historically these residues were first identified either by evaluating the sequence conservation among the class A GPCR family or by comparing the two structures of GPCR subtype in different states; and the functional importance of the selected residues was subsequently confirmed by mutagenesis studies.[1,20] Thus, a receptor belonging to the class A GPCRs is expected to utilize many of these 18 microswitches for allosteric signaling. Although one should still be mindful of the fact that the functional role of these microswitches have not been verified for all the GPCR subtypes, the 18 microswitches can be used as benchmark residues to assess the performance of a prediction tool on allosteric hotspots in GPCRs (see Materials and Methods). The extent of sequence conservation in class A GPCRs, quantified by evaluating the sequence conservation free energy ΔG [Eq. (1)] indicates that 15 out of 18 microswitches (except for P189, S281, and E228) are highly conserved, satisfying $\Delta G(\text{GPCR})/k_B T^* \geq 0.2$ [Fig. 3(A,B)]. Figure 3(C) visualizes how rotameric transition is made from the inactive to active state and highlights the difference in the orientation of the side chain in some of the microswitches by contrasting the apo and agonist-bound states.

The allosteric hotspots of A_{2A} adenosine receptor mediate the flow of information

As a tool for studying protein allostery, the network centrality, a measure that quantifies the degree of centralization of a node in network theories, can be employed to unravel the hotspot residues of a given protein network. Among the popular centrality measures in network theories[21] (degree (C_D), closeness (C_C), and betweenness (C_B) centralities, whose definitions are given in Materials and Methods), the betweenness centrality, $C_B(v)$, evaluates the extent to which the node v has control over the information flow in the network.[22] Conceptually, it could be argued that a node of high C_B value is the spot mediating the allosteric signal flow (Materials and Methods). By using the minimum energy structures obtained from MD simulations we constructed the residue interaction network for the apo and agonist-bound states of the A_{2A}AR by taking into account the presence of side chains [Fig. 4(A), see Materials and Methods], and calculated

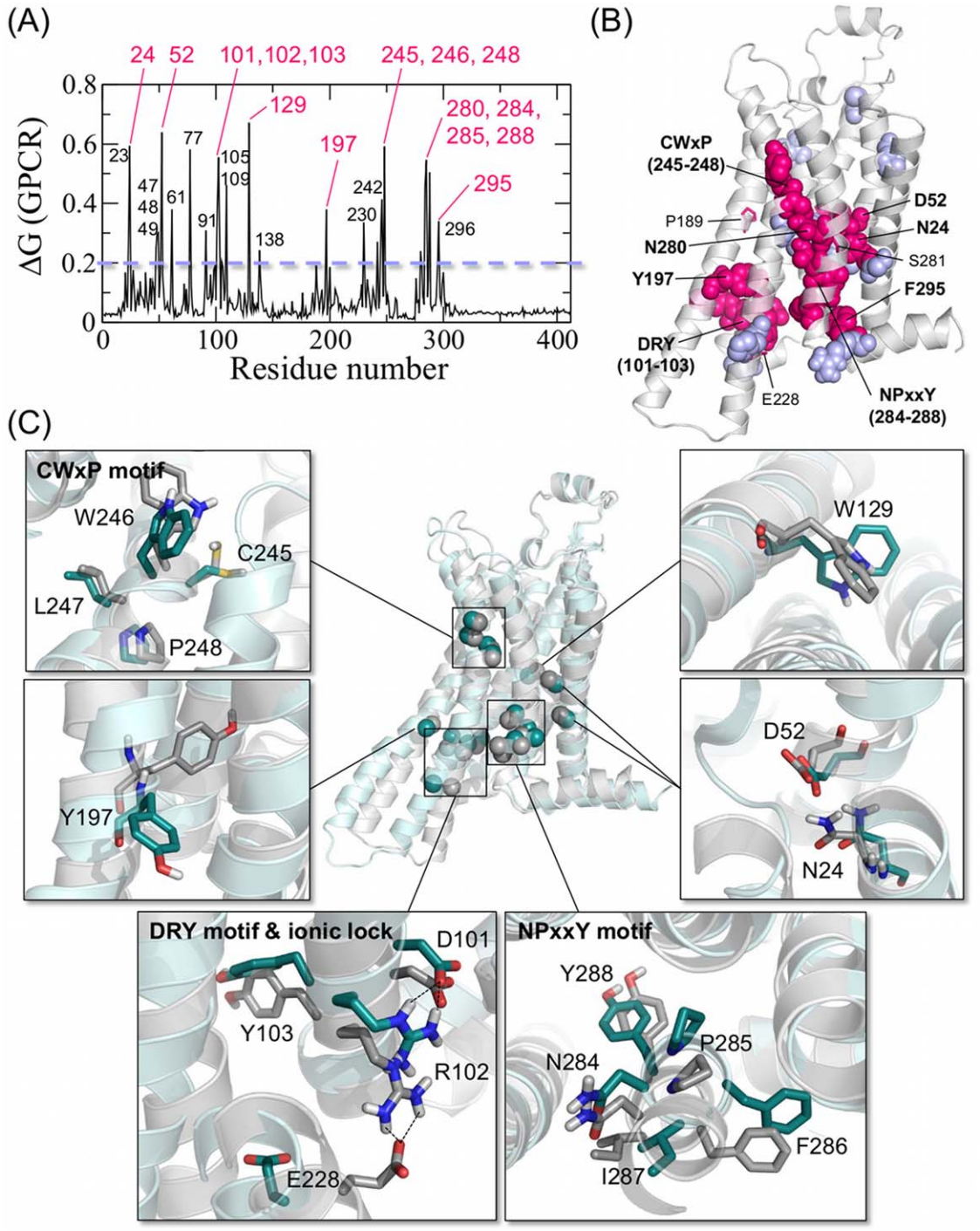
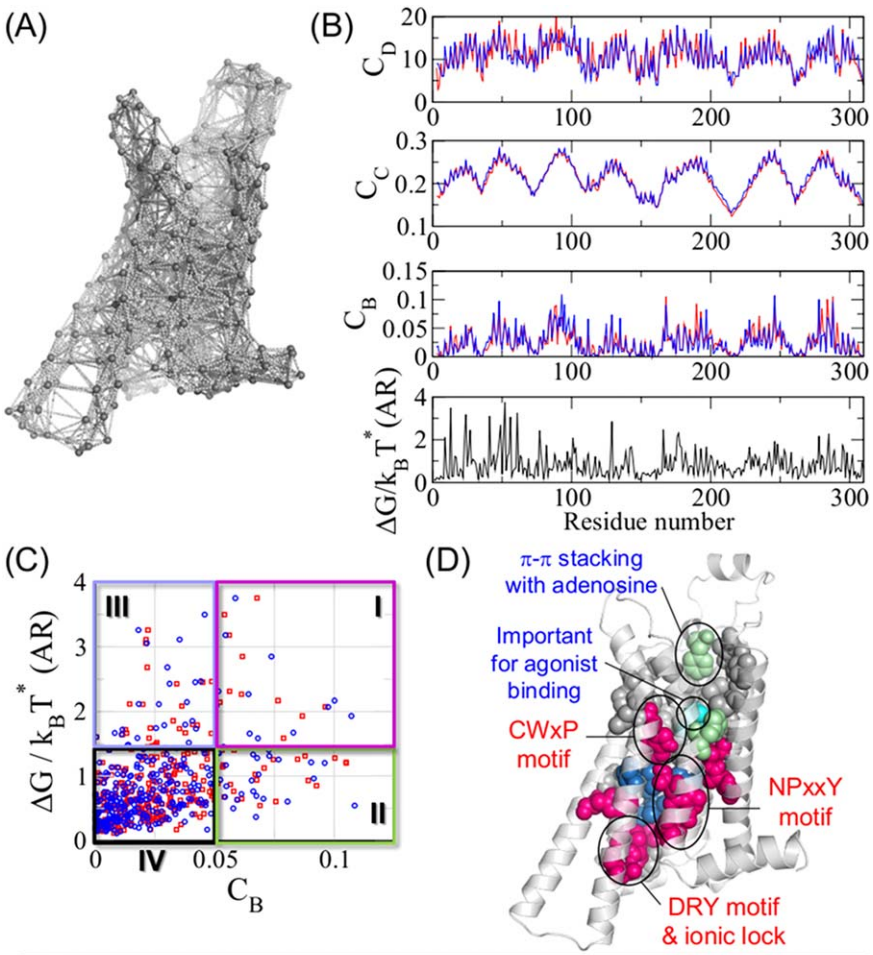


Figure 3

Microswitches in GPCRs. (**A**) Sequence conservation free energy ($\Delta G/k_B T^*$) computed for the class A GPCR family. The specification of GPCR inside the parenthesis of ΔG(GPCR) in the figure indicates that $\Delta G/k_B T^*$ value was calculated for a multiple sequence alignment for the class A GPCR family. The residues with $\Delta G \geq 0.2$ are annotated; and among them 15 residues identified as microswitches in literatures are highlighted in magenta. (**B**) Residues of $\Delta G/k_B T^* \geq 0.2$ are depicted with spheres on $A_{2A}AR$ structure. Among them, microswitches are colored in magenta, and others with $\Delta G/k_B T^* > 0.2$ are in light-blue. The residues P189, E228, and S281, that are proposed as microswitches in literatures but have $\Delta G/k_B T^*$ less than 0.2, are depicted using stick representation. (**C**) Conformational changes of the key structural motifs and microswitch residues are depicted using the minimal energy structures of apo (grey) and agonist-bound forms (cyan) obtained from our MD simulations. It is proposed that the rotameric transitions of microswitches are critical for the intra-molecular signal transmission of GPCRs.



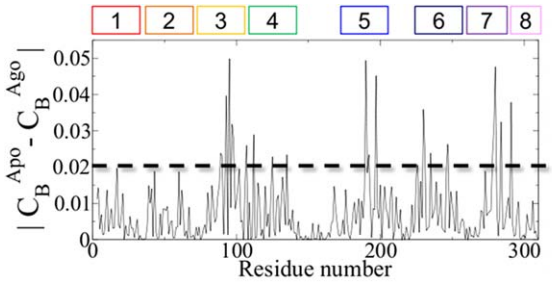
I	N24, *L48, D52, D101, W129, *M177, M193, Y197, E228, W246, *H278, P285
II	F44, V55, I60, I64, I66, *V84, L85, L87, *T88, *Q89, *S91, I92, F93, L95, A97, I98, R102, Y112, I125, I135, F168, F182, *V239, L247, *H250, N284, Y288
III	Y9, E13, G23, V27, T41, A49, A51, G56, P61, C77, C82, Y103, G142, W143, C166, Y176, V178, Y179, P189, P248, N253, F295
IV	others (C245, N280, S281, …)

†The residues with $|C_B^{Ago} - C_B^{Apo}| > 0.02$ are underlined.
*Mutation data is available in GPCRDB.

Figure 4

Network centrality analysis of A_{2A} adenosine receptor. (**A**) Network representation of the A_{2A}AR in apo form built by taking into account the presence of side chain in each residue (see Materials and Methods). (**B**) Degree (C_D), closeness (C_C), betweenness centralities (C_B) for each residue of the A_{2A}AR (blue: the apo form; red: the agonist-bound form) and the sequence conservation free energy ($\Delta G/k_BT^*$) calculated for AR family. (C) Scatter plot of ($\Delta G/k_BT^*$, C_B) (blue: apo form; red: agonist-bound form). Based on $C_B(i)$ and $\Delta G_i/k_BT^*$ values, the residues of A_{2A}AR were categorized into four groups from I to IV. The residues with high $C_B(\geq 0.05)$ (group I and II) and with high sequence conservation ($\Delta G/k_BT^* \geq 1.5$) in AR family (group I and III) are depicted on the apo structure of the A_{2A}AR in Supporting Information Figure S2A and S2B, respectively. (**D**) Among the residues that belong to the groups I and II with $C_B^{Apo} \geq 0.05$, key residues confirmed from the previous biochemical studies for class A GPCRs are marked on the A_{2A}AR structure using different colors (magenta for the microswitches (see Supporting Information Fig. S3 for the top and bottom views): cyan for the residue important for agonistic binding; pale green for the residues important for ligand binding; blue for the residues in hydrophobic barrier); the underlined residues satisfy the condition $|C_B^{Apo} - C_B^{Ago}| \geq 0.02$; and the residues marked with asterisks are those whose mutation data is available in GPCRDB for A_{2A}AR.

$C_D(v)$, $C_C(v)$, and $C_B(v)$ [Fig. 4(B)]. The overall correlations between the different centrality measures are not that strong (correlation coefficient = 0.66−0.72) (Supporting Information Fig. S1); thus a residue with high C_D (or equivalently with a large number of contacts) or C_C does not necessarily retain a high C_B value. Among the three network centralities, C_B exhibits the highest selectivity [Fig. 4(B)]. As depicted on the A_{2A}AR structure, the residues with $C_B \geq 0.05$ (the top 10% of the C_B-distribution), which are deemed important for



$\lvert C_B^{Apo} - C_B^{Ago} \rvert \geq 0.02$	
Group I	197(5), 246(6), 278(7)
Group II	89(3), 93(3), 95(3), 97(3), 98(3), 112(3), 125(4), 135(4), 247(6), 284(7)
Group IV	90(3), 107(3), 190(5), 192(5), 226(6), 230(6), 231(6), 235(6), 279(7), 280(7), 291(8)

†Number in parenthesis indicates the TM number.

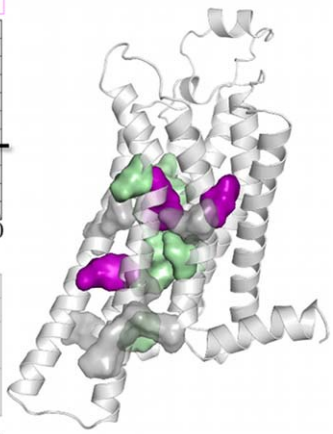


Figure 5

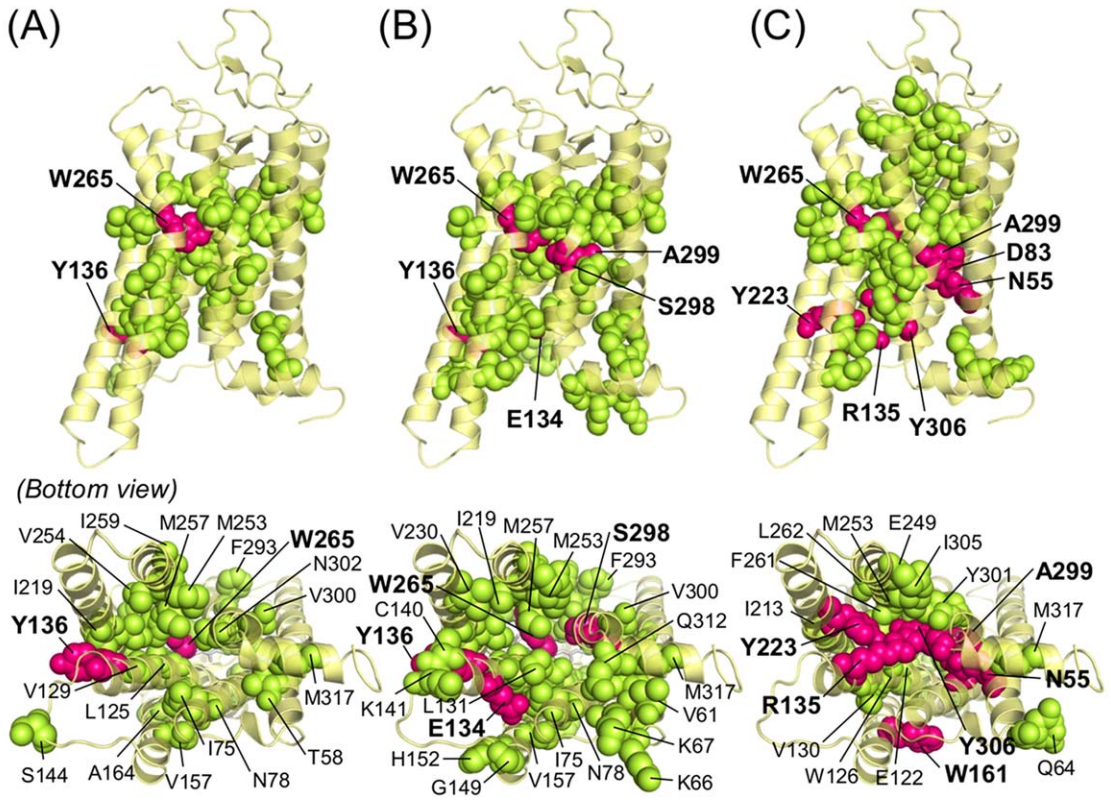
Difference of C_B values calculated for apo and agonist-bound structures. Residues with $|C_B^{Ago} - C_B^{Apo}| \geq 0.02$, contributed from TM3, 5, 6, and 7, are depicted with magenta for group I, green for group II, grey for group IV, and their indices are listed on the table.

controlling information flow from the definition of C_B, are distributed contiguously, bridging the extracellular (EC) and intracellular (IC) parts of TM helices [Fig. 4(D)].

The 39 allosteric hotspots of A_{2A}AR predicted using $C_B \geq 0.05$ [the residues are listed in the groups I and II in Fig. 4(D)] include many important residues suggested from biochemical studies for class A GPCRs in general and A_{2A}AR in particular. Among the 18 residues suggested as general microswitches for class A GPCRs, 11 of them (N24, D52, D101, R102, W129, Y197, E228, W246, N284, P285, Y288) are identified by the simple condition of $C_B \geq 0.05$. Since the probability of correctly identifying at least a single microswitch from random drawing is $p_m = 18/308 \approx 0.06$, the expectation value of identifying microswitches by selecting 39 residues is $39 \times p_m \approx 2.3$ (see Materials and Methods). Given that we identified as many as 11 residues the performance of our C_B-based analysis should be considered significant. Among the residues identified by this condition other than microswitches, $F44^{2.42}$, $L48^{2.46}$, $L95^{3.43}$, $I98^{3.46}$, and $V239^{6.41}$ [blue in Fig. 4(D)] compose a region called the hydrophobic barrier that separates CWxP and NPxxY motifs from DRY motif;[56] F168 in ECL2, $H278^{7.43}$, and $T88^{3.36}$ [green in Fig. 4(D)] are the residues known to be important for ligand binding in AR family.[26] F168 can potentially interact with adenine ring of nucleoside ligands via π-π stacking. $T88^{3.36}$ in the TM3 helix that can form a hydrogen bonding with an agonist is important for sensing the agonist binding and transmitting signals to the intracellular G-protein binding site;[26,57,58] L48, M177, V84, T88, Q89, S91, H250 [marked with asterisks in Fig. 4(D)] are also found essential for receptor function of A_{2A}AR according to the mutation data in GPCRDB.[59] Lastly the residues identified by $C_B \geq 0.05$ but not commented above to have any overlap with the previous biochemical studies (M193, V55, I60, I64, I66, L85, L87, I92, F93, A97, Y112, I125, I135, F182, L247) could be regarded as candidate residues for allosteric hotspots of A_{2A}AR that our C_B-based analysis predicts, which are amenable to further experimental study.

In conjunction with C_B value, the extent of sequence conservation in each residue, $\Delta G(\mathrm{AR})/k_BT^*$ [Eq. (1)], based on the multiple sequence alignment of adenosine receptor subfamily, could be useful for the purpose of our analysis. Here, it should be noted that $\Delta G(\mathrm{AR})$ is different from $\Delta G(\mathrm{GPCR})$ in Figure 3(A). $\Delta G(\mathrm{AR})$ is calculated by restricting the MSA to the subfamily of adenosine receptors while $\Delta G(\mathrm{GPCR})$ is calculated using the entire MSA for class A GPCRs. Partitioning the residues into four different groups based on the $\Delta G/k_BT^*$ and C_B scores [Fig. 4(C)], i.e., $C_B \geq 0.05$, $\Delta G/k_BT^* \geq 1.5$ for group I; $C_B \geq 0.05$, $\Delta G/k_BT^* < 1.5$ for group II; $C_B < 0.05$, $\Delta G/k_BT^* \geq 1.5$ for group III; $C_B < 0.05$, $\Delta G/k_BT^* < 1.5$ for group IV, we make a few points below.

a. First, the definitions of C_B and $\Delta G/k_BT^*$ are totally independent from each other. Evident from the scatter plot in Figure 4(C), no clear correlation is found between C_B and $\Delta G(\mathrm{AR})/k_BT^*$. Yet, the commonly identified residues with the conditions of high C_B (≥ 0.05) and high $\Delta G/k_BT^*$ (≥ 1.5), namely the group I residues contain as many as eight microswitches and four other hotspot residues. The group I residues ($C_B \geq 0.05$, $\Delta G/k_BT^* \geq 1.5$) are clustered at the core region of TMs (magenta region in Supporting Information Figs. S2A and S2B), the contiguous surface of which is known to form hydrogen-bond network with the conserved polar residues and structural water molecules.[20,60] Of particular note is that evolutionarily covarying residues identified from the



	†Identified functionally important residues
SCA by *Süel et al.* (*Nat. Struct. Biol.* **2003**, *10*, 59; Figure 4)	58, 75, 78, 91, 92, 113, 122, 123, 125, 129, ***136,** 144, 157, 164, 170, 212, 213, 219, 222, 253, 254, 257, 259, 261, ***265,** 268, 293, 296, 300, 302, 317
SCA by *Dima et al.* (*Protein Sci.* **2006**, *15*, 258; Table 1)	44, 51, 58, 59, 61, 66, 67, 75, 78, 90, 91, 92, 113, 115, 117, 120, 121, 122, 123, 124, 125, 126, 129, 131, ***134,** ***136**, 140, 141, 149, 152, 157, 164, 168, 213, 219, 222, 230, 249, 253, 254, 257, 258, 259, ***265,** 268, 269, 293, 294, 295, 296, ***298, *299,** 300, 312, 317
Betweenness centrality analysis ($C_B \geq 0.05$)	12, ***55,** 64, ***83,** 86, 91, 113, 122, 125, 126, 128, 130, ***135, *161,** 178, 179, 181, 184, 186, 189, 192, 207, 208, 213, ***223,** 249, 253, 261, 262, 264, ***265,** 268, 296, ***299,** 301, 305, ***306,** 317

†The residues are numbered based on the bovine rhodopsin.
*The microswitches are marked with asterisks.

Figure 6

Comparison between the allosteric hotspots for rhodopsin predicted by SCA and CB-based analysis. Hotspots identified from SCA by (**A**) Suel et al.,[12] and (**B**) Dima et al.,[13] and (C) from our network analysis based on the residues with $C_B \geq 0.05$. Each method detected (A) 2 (B) 5 (C) 8 microswitches.

statistical coupling analysis (SCA), by definition, cannot have a high $\Delta G/k_BT^*$ value; thus SCA cannot detect residues in group I. The efficacy of C_B score in identifying microswitches as well as other hotspots is compared with SCA for the case of rhodopsin in the following section and Figure 6.

b. For AR family, most of the residues with low C_B but with high ΔG score (residues belonging to the group III) are distributed around the ligand binding site and in the cytoplasmic side (Supporting Information Fig. S2B). The high sequence conservation in the ligand binding sites, identified from the MSA of AR subfamily, is consistent with our general notion that adenosine receptors are specific to the adenosine ligand, which allows the receptor to effectively discriminate other ligand types. On the other hand, when MSA is carried out for the entire sequences of the class A GPCR family, the conserved residues are identified more at the cytoplasmic region where G-protein binds [Fig. 3(B)]. These findings suggest that the subtype specificity or functional classification is correctly captured in residues with high ΔG value as long as a good MSA is used.

c. There are slight differences in the C_B scores between the apo and agonist-bound forms. The contribution of residues satisfying the condition $|C_B^{Ago} - C_B^{Apo}| \geq$

0.02 comes from the group I (12.5 %, residue number: 197, 246, 278), group II (37.5 %, residue number: 89, 93, 95, 97, 98, 112, 125, 135, 247), and group IV (50 %, residue number: 90, 107, 190, 192, 226, 230, 231, 235, 279, 280, 284, 291) (Fig. 5). The residues identified with high $|C_B^{Ago} - C_B^{Apo}| \geq 0.02$ values are mainly located in TM3 and TM5–7. Of particular note is that majority of the residues with $|C_B^{Ago} - C_B^{Apo}| \geq 0.02$, also satisfying $C_B^{Apo} \geq 0.05$, are found in the group II [9 out of 12, these residues are marked with underlines in the table of Fig. 4(D)], which suggests that among the allosteric hotspots (groups I and II) the less conserved residues (group II) are more sensitive to the *apo* → *ago* (or inactive → active) conformational change.

As presented above, the C_B-based network analysis of the A_{2A}AR structure enables us to identify the allosteric hotspots of A_{2A}AR that show neither the sequence variation nor a detectable conformational change in the transition from the apo to agonist-bound form. Next, we will show that the performance of C_B-based analysis in identifying the location of microswitches is remarkable by making quantitative comparisons with other conventional approaches.

Comparison with other approaches

Statistical coupling analysis (SCA)

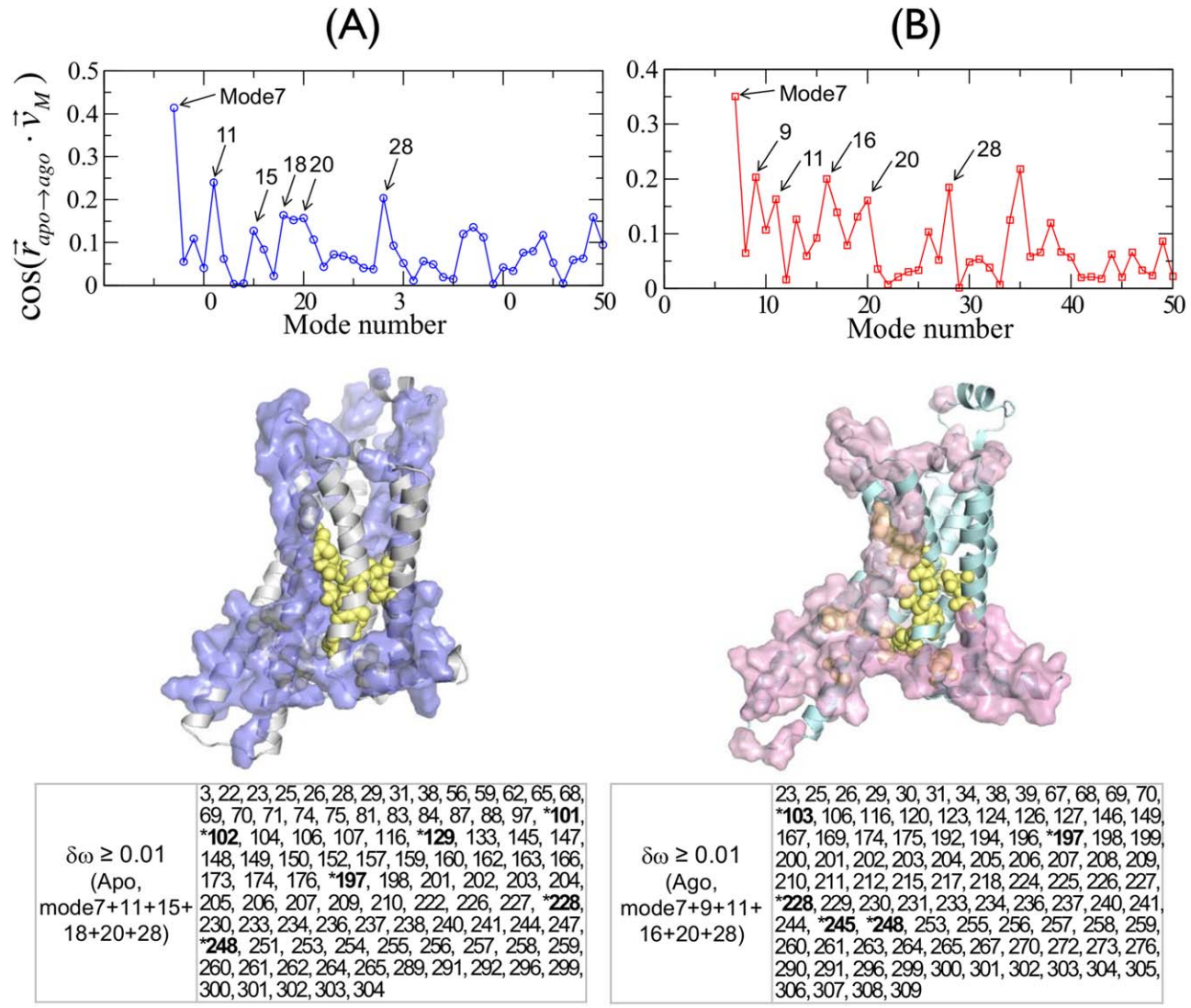
A strong signal of covariation between two remote residues in a multiple sequence alignment, which is exploited as a basic principle to identify clusters of residues under long-range coupling in a bioinformatical method called statistical coupling analysis (SCA),[11–14] is viewed as a consequence of allosteric communication mediated by multiple groups of residues that lie in the midst of signaling pathways. While it was proposed that the method using SCA on GPCR identified the "sparse network of coevolving amino acids" (or sectors)[61,62] that bridges the ligand-binding site to the cytoplasmic G-protein interaction site, forming the allosteric signaling pathways,[12,13] it fails to detect several highly conserved microswitches. Figure 6 shows the list of allosteric hotspots identified for bovine rhodopsin by SCA from two different studies [Figs. 6(A,B)] and the residues with high $C_B(\geq 0.05)$ [Fig. 6(C)]. Although the two methods are based on entirely different assumptions, one solely based on sequence information, the other on network topology, allosteric hotspots identified for rhodopsin are mainly distributed around the TM region. It should, however, be noted that C_B-based network analysis is much more efficacious in identifying the microswitches, which are considered critical in the activation mechanism of class A GPCRs. For rhodopsin, SCA using two slightly different definitions of $\Delta G/k_B T^*$ in Refs. 12 and 13 identifies two and five microswitches, respectively, whereas our C_B-based analysis identifies eight microswitches out of 18 predicted residues.

Statistical assessment of three results in Figure 6 can be made by calculating φ_m, the ratio between the number of predicted microswitches (N_m) and the expectation value ($\langle n \rangle_{rand}$) [Eq. (2)]. The number of correctly identified microswitches (N_m) and the number of residues selected for the prediction (n) in each method are $(N_m, n) = (2, 31)$, (5,55), and (8,38) for (i) Suel *et al.*, (ii) Dima *et al.*, and (iii) C_B-based analysis, respectively. Therefore, $\varphi_m^{(i)} = 2/1.8$, $\varphi_m^{(ii)} = 5/3.2$, and $\varphi_m^{(iii)} = 8/2.2$. Note that $(\varphi_m^{(i)} < \varphi_m^{(ii)} < \varphi_m^{(iii)})$ indicates that prediction of microswitches made by C_B-based analysis is better than those made by SCA, attesting to the utility of C_B-based analysis.

Structural perturbation method (SPM)

The SPM is used to identify key residues controlling the conformational dynamics by assessing the importance of a residue in the elastic network representation under local perturbation.[15,16] The perturbation is invoked by changing the force constant of the springs that link the residue and its neighbors. When the overlap of mode M ($\vec{v}_M$) with the vector defining the transition of apo to agonist-bound form ($\vec{r}_{apo \to ago} = \vec{R}_{ago} - \vec{R}_{apo}$) is significant, i.e., when $\cos(\vec{r}_{apo \to ago} \cdot \vec{v}_M)$ is large, the frequency change of a mode M under the perturbation of i-th residue is calculated using $\delta\omega(M, i) = \vec{v}_M^T \cdot \delta H \cdot \vec{v}_M$, where δH is the Hessian matrix of the following perturbed energy potential for elastic network model: $\delta E_{ENM} = \frac{1}{2} \sum_{ij} \delta k_o (r_{ij} - r_{ij}^o)^2 \Theta(r_{ij}^o - R_c)$. Note that the expression of $\delta\omega(M, i) = \vec{v}_M^T \cdot \delta H \cdot \vec{v}_M$ is analogous to the first-order energy correction term for the M-th eigenmode in non-degenerate perturbation theory.[15,16] Thus, if a perturbation on the i-th residue leads to a large change in $\delta\omega(M, i)$, the residue i is considered to be important in the SPM.

We found that in both for apo and agonist-bound structure the mode 7 (excluding the three translational and three rotational modes, the mode 7 is the lowest eigenmode) has maximum overlap with the conformational change $\vec{r}_{apo \to ago}$ (Fig. 7). As shown in Figure 7, key residues with high $\delta\omega$ are mainly distributed in the extracellular and intracellular regions of TM helices, which are accompanied with large conformational changes when the transition occurs from the apo to agonist-bound form. Note, however, that even the superposition of six major modes, which have large overlap with conformational changes, is not good enough to identify microswitches that are buried deep inside the GPCR structure. Prediction efficiency of SPM that identifies $N_m = 6$ and 5 microswitches out of $n = 98$ and 97 residues for apo and agonist-bound forms (Fig. 7) is only $\varphi_m = 1.1$ and 0.9, respectively [see Eq. (2)]. Whereas, $\varphi_m = 4.8$ for C_B-based analysis indicates that C_B-based



δω ≥ 0.01 (Apo, mode7+11+15+18+20+28)	3, 22, 23, 25, 26, 28, 29, 31, 38, 56, 59, 62, 65, 68, 69, 70, 71, 74, 75, 81, 83, 84, 87, 88, 97, ***101**, ***102**, 104, 106, 107, 116, ***129**, 133, 145, 147, 148, 149, 150, 152, 157, 159, 160, 162, 163, 166, 173, 174, 176, ***197**, 198, 201, 202, 203, 204, 205, 206, 207, 209, 210, 222, 226, 227, ***228**, 230, 233, 234, 236, 237, 238, 240, 241, 244, 247, ***248**, 251, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 264, 265, 289, 291, 292, 296, 299, 300, 301, 302, 303, 304
δω ≥ 0.01 (Ago, mode7+9+11+16+20+28)	23, 25, 26, 29, 30, 31, 34, 38, 39, 67, 68, 69, 70, ***103**, 106, 116, 120, 123, 124, 126, 127, 146, 149, 167, 169, 174, 175, 192, 194, 196, ***197**, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 215, 217, 218, 224, 225, 226, 227, ***228**, 229, 230, 231, 233, 234, 236, 237, 240, 241, 244, ***245**, ***248**, 253, 255, 256, 257, 258, 259, 260, 261, 263, 264, 265, 267, 270, 272, 273, 276, 290, 291, 296, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309

Figure 7

SPM-identified residues with high δω values for the superposition of high overlapping modes (M=7−28) of (**A**) the apo (blue) and (**B**) the agonist-bound form (red). The degree of overlap, $\cos(\vec{r}_{apo\to ago}\cdot\vec{v}_M)$, calculated between the conformational change from apo to agonist-bound state and Mth normal mode (top). The superposition of hotspot residues, satisfying $\delta\omega(M,i)\geq 0.01$ for high overlapping modes, are depicted with blue and magenta surfaces, respectively, and their residue numbers are listed in the table below, in which the microswitches are marked with asterisks. Note that the key residues identified by SPM are mainly located around the hinge region controlling the motion of TM5-ICL3-TM6. For comparison, the locations of the microswitch residues are depicted with yellow spheres. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

analysis certainly outperforms SPM in identifying microswitches. Hence, neither is the SPM suitable for identifying the microswitches of GPCRs, which undergo only a minor change in their positions before and after the activation.

The microswitches are critical for the integrity of signaling network of GPCRs.

In theory of complex networks, a network's tolerance to an error or vulnerability to an attack is evaluated using the relative change in the average network centrality when a node, say x, is removed,[33] which can be written as follows:

$$\Gamma_\xi^x=\frac{\langle C_\xi\rangle-\langle C_\xi^x\rangle}{\langle C_\xi\rangle} \tag{6}$$

where $\langle C_\xi\rangle(\equiv\sum_{i=1}^{N}C_\xi(i)/N)$ is the average network centrality, and $\langle C_\xi^x\rangle$ is a value evaluated for a newly constructed network when the node x is removed from the original network. The idea of network vulnerability is, in fact, routinely practiced in molecular biology in the form of protein mutagenesis assay, which measures the effect

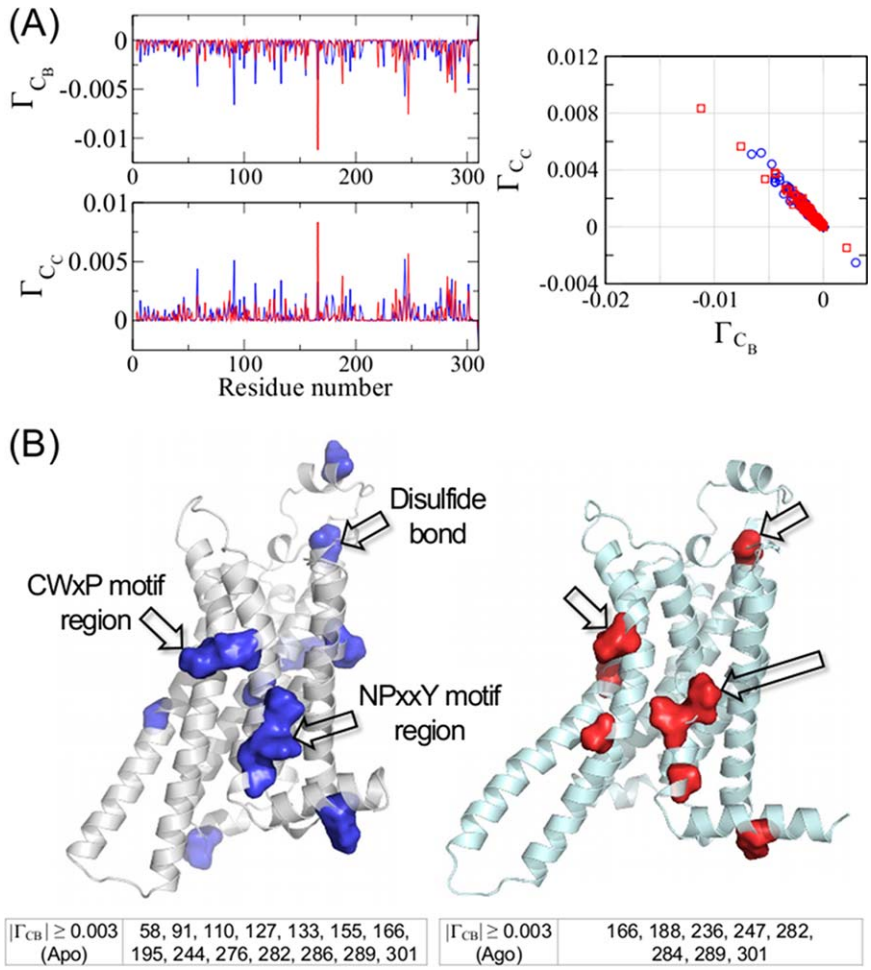


Figure 8

Glycine scanning (network vulnerability) of A_{2A} adenosine receptor. (**A**) Γ_{C_B} (top) and Γ_{C_C} (bottom) (blue: the apo structure; red: the agonist-bound structure). Scatter plot of (Γ_{C_B}, Γ_{C_C}) is shown to indicate that Γ_{C_B} and Γ_{C_C} are well correlated. (**B**) Regions with high network vulnerability ($|\Gamma| \geq 0.003$) in the apo (left) and agonist-bound forms (right) are represented with blue and red surfaces, and corresponding residue indices are listed in the table.

of mutations on the degree to which proteins can retain their activity. Adapting the idea of network vulnerability, we performed in silico glycine scanning of the constructed residue interaction network of the A_{2A}AR. As straightforwardly implicated by the term "glycine scanning", we mimicked the protein mutagenesis assay by deleting the side chain of each residue and evaluated the deletion effect on the network. Our glycine scanning analysis differs from the previous study applying network analysis[24] in that only a side chain, rather than the entire residue, is deleted for each scan. It is important to keep Cα backbone because, even in the absence of the side chain, intra-molecular residue contacts can still be formed via backbone-side chain or backbone-backbone interactions. Note that here a readjustment of local environment due to the side chain removal is not considered. Our aim here is to make a quantitative assessment of the role of the side chain in the original residue interaction network. The greater is the role played by the removed side chain in maintaining the network structure, the more significant would be the response of average network centrality to the removal of that particular residue. We assess the effect of deleting side chains by calculating the changes in average closeness ($\langle C_C \rangle$) or betweenness centralities ($\langle C_B \rangle$), both of which turn out to be highly correlated [Fig. 8(A)].

Our glycine scanning analysis identified the group of residues critical for the integrity of interaction network that is responsible for the receptor allostery. The residues with strong network vulnerability ($|\Gamma_{C_\xi}| \geq 0.003$) are identified in the regions around CWxP and NPxxY motifs [Fig. 8(B)],[27] which retain proline that creates a kinked helix in the middle of TM6 or TM7.[20] In the inactive state of GPCRs, interactions between the

cytoplasmic ends of TM3 and TM6 constrain the relative motion of these segments by forming an ionic-lock between R102$^{3.50}$ and E228$^{6.30}$.[20] Disruption of such constraint, triggered by agonist binding, enables TM6 to move outward from TM3 [see DRY motif and ionic lock in the Fig. 3(C)]. NPxxY motif, which interacts with TM6 or helix 8, imposes structural constraints in GPCRs and stabilizes the helical structures.[63,64] In addition, C166, which constrains ECL1 and ECL2 by forming a disulfide bond with C77, is detected to have high network vulnerability. It is of note that the constrained random coil structure of ECL2 is unique to A_{2A}AR in that the ECL2 of other GPCRs typically forms β sheet or α-helix.[27]

Distinct C_B-based wiring diagrams reflect GPCR subtype specificity

Here we extend the C_B-based network analysis to other class A GPCRs, including β_1, β_2 adrenergic receptors (PDB IDs: 2VT4 and 3NYA), chemokine CXCR4 receptor (3ODU), dopamine D3 receptor (3PBL), histamine H1 receptor (3RZE), and bovine rhodopsin (1U19).[65] Similar to the A_{2A}AR, the network of residues with high $C_B(\geq 0.05)$ in these class A GPCRs form contiguous surface that bridges between the ligand binding and G-protein binding sites (Supporting Information Fig. S4). In most GPCRs the high C_B residues are mainly distributed around the "minor binding pocket," located in the shallow part of the ligand binding site between the TM1, 2, 3 and TM7, which serves as an onset point of orthosteric signal transduction process.[66] In particular, when the C_B is restricted to a value greater than 0.075, the high C_B residues bridge the extracellular region of TM3 to the TM6, 7 helices. For the class A GPCRs, the highly vulnerable residues identified by the glycine scanning analysis are mostly distributed in TM3 and TM7 (Supporting Information Fig. S5). Note that K296$^{7.43}$ in bovine rhodopsin, known to contribute to the activation of rhodopsin by forming a covalent bond with retinal,[66] is also identified (the residue marked with a yellow arrow in Supporting Information Fig. S5G). Along with the variation of residues (F168 in A_{2A}AR; R183 and Y190 in CXCR4 receptor; and K179 in H1 receptor) and wiring diagram in ECL2 detected by glycine scanning analysis, the variations in the high C_B-surfaces demonstrated in the class A GPCRs (Supporting Information Fig. S4) are deemed responsible for their subtype selectivity.[66]

Long-range transmembrane cross-correlation in the agonist-bound active state

As suggested by the G-protein bound structure,[1] it is expected that the agonist binding site in the extracellular side and intracellular region are functionally coupled in the active forms of GPCRs, and this coupling is mediated by a structural reorganization of seven membered TM helices. To quantify such long-range coupling in the dynamics of A_{2A}AR, we calculated cross-correlation between residues in terms of C_B [Eq. (2)] by using the conformational ensemble of the A_{2A}AR generated from the 300 nsec MD simulation trajectory (see Materials and Methods and Supporting Information Fig. S9).

$$CC_{ij} = \frac{\langle C_B(i)C_B(j)\rangle - \langle C_B(i)\rangle\langle C_B(j)\rangle}{\sqrt{\langle(\delta C_B(i))^2\rangle}\sqrt{\langle(\delta C_B(j))^2\rangle}} \qquad (7)$$

where $\langle \ldots \rangle$ refers to an ensemble average; thus $\langle C_B(i)\rangle$ denotes the average betweenness centrality for the residue *i*. As shown in the cross-correlation matrices for apo and agonist-bound form (Supporting Information Figs. S6 and S7), the signatures of correlation between residues are scattered all over the structure. To identify residue pairs with long-range cross-correlation we imposed the conditions of $CC_{ij} \geq 0.5$ and $d_{ij} > 6$ [Fig. 9(A)]. Importantly, while in the apo structure the residue pairs under high cross-correlations are distributed only around the cytoplasmic side of the TMs [Fig. 9(A) and Supporting Information Fig. S7), functionally important long-range couplings are detected between the ligand-binding and cytoplasmic G-protein binding sites in the agonist-bound form [Fig. 9(A) and Supporting Information Fig. S7]. This result from the agonist-bound form is consistent with the view that a bound agonist makes tight interactions with the surrounding residues and increases the receptor activity above its basal level.[1] The long-range coupling between the ligand binding site and G-protein binding site for the agonist-bound form is also grasped by computing the mean square fluctuation using structural ensemble (see Supporting Information Fig. S8).

Notably, there are multiple parallel paths linking the correlated residues,[67] the degeneracy of which varies from 1 to as many as 480 depending on the residue pair (For the details of entire paths between the correlated residues, see Part 1 and 2 in Supporting Information II). The presence of multiple parallel pathways is consistent with the recent new view of allostery.[50,67,68] As some of the representative allosteric paths, linking the residues in extracellular and intracellular regions, are demonstrated in Figure 9(B), the 80 % of transmembrane signaling paths go through the residues with high C_B, which includes the microswitches as well as other functionally important residues (see the residues represented with cyan spheres in Figure 9(B,C) and Part 1, 2 in Supporting Information II). It is these residues, lying in the midst of communication pathways, that toggle the intramolecular signaling. The qualitatively disparate results displayed in apo and agonist-bound forms provide a picture consistent with the function of GPCRs.

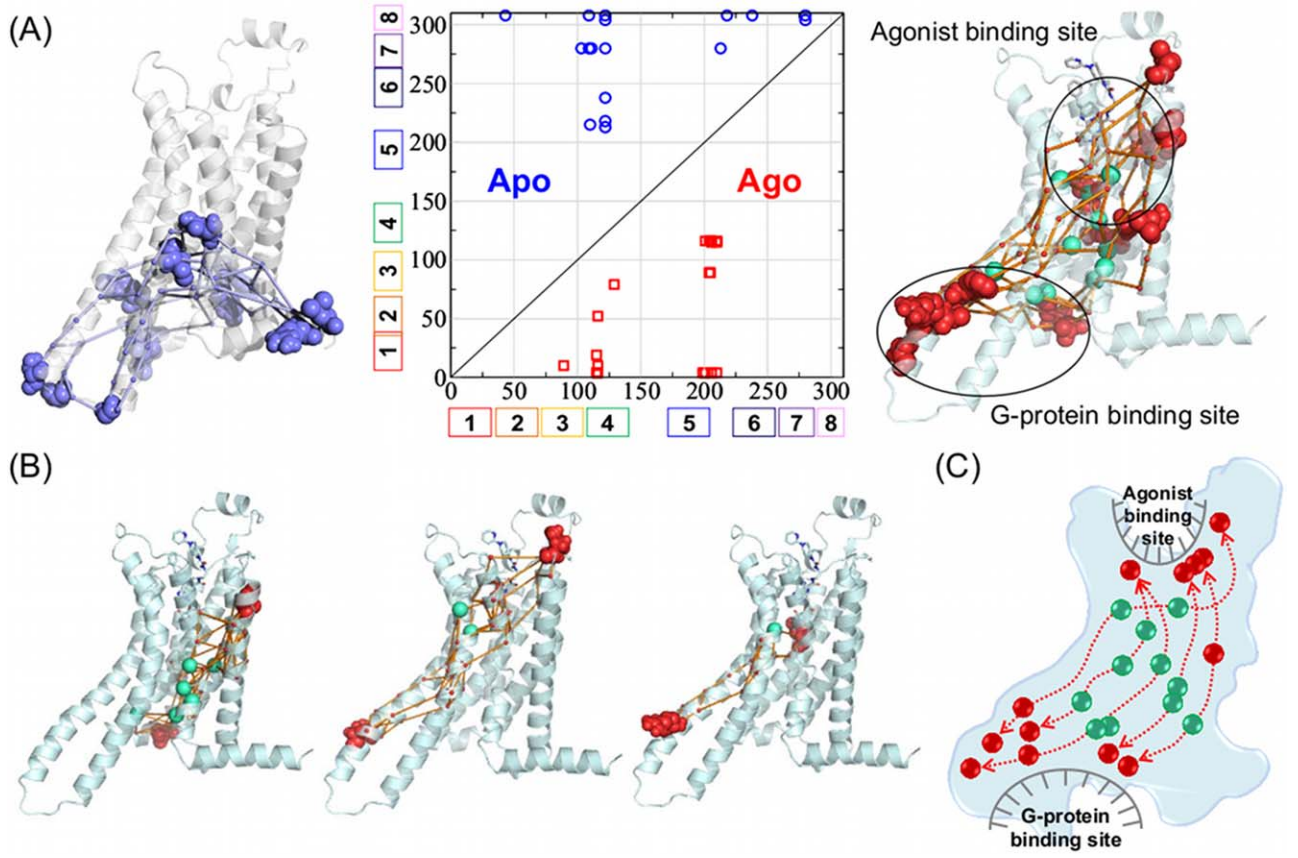


Figure 9

Multiple pathways of intramolecular signals that link the cross-correlated residues in the extra and intra-cellular domains of $A_{2A}AR$. (**A**) Residue pairs with high cross-correlation ($|CC_{ij}| \geq 0.5$) and distance greater than 6 ($d_{ij} > 6$) are marked using blue circles and red squares for the apo form (upper left corner) and for the agonist-bound form (lower right corner), respectively (see the original map in Supporting Information Fig. S6). The minimum paths between the cross-correlated residues are shown for apo and agonist-bound forms on the left and right, respectively. For agonist-bound form, long-range cross-correlations are detected between the extracellular ligand binding and cytoplasmic G protein binding sites. Microswitches and other residues with high C_B on the paths are displayed in cyan spheres. (**B**) Examples of the multiple signaling paths between the agonist binding and G-protein binding sites. (**C**) Schematic of transmembrane signaling represented by multiple shortest paths linking the long-range cross-correlated residues.

To systematically group correlated residues, we carried out hierarchical clustering analysis on the acquired matrices and represented the result using dendrogram (Supporting Information Fig. S7). From the two clusters of positively correlated residues (cluster 1 and 2), the clusters of residue pairs with the strongest signal are shown on each clustered cross-correlation map of the apo and agonist-bound form with residue indices. The cross-correlated residue pairs obtained using hierarchical clustering analysis (Supporting Information Fig. S7) are similar to those from the simple condition of $CC_{ij} \geq 0.5$ and $d_{ij} > 6$ that we imposed in Figure 9. The residues within each of cluster 1 and 2 are the parts of structure that "breathe together" in terms of C_B values. Also, it is noteworthy that in terms of the correlation of C_B value there is a strong anti-correlation between cluster 1 and cluster 2, which suggests that "breathing" of residues in cluster 1 and cluster 2 occurs out-of-phase.

Lastly, it is worth considering the signaling paths on a weighted graph. To this end, we employed the "dynamical network community analysis,"[43] implemented to the molecular visualization package VMD. In this analysis, an inter-residue cross-correlation calculated from an ensemble of structures from long MD simulation is used for the weight of edges in the network. Using the NetworkView module in VMD and our 300 ns MD simulation as an input, we calculated the optimal and a set of suboptimal paths (offset = 20) between residue pairs that show long-range cross-correlation (Supporting Information II—Part 3). In most of the cases, the allosteric signaling paths computed on unweighted and weighted graphs for agonist-bound form are qualitatively similar; yet, it is interesting to point out the large detour in the signaling paths of the residue pairs 116-4 and 116-10 on weighted graph (see Fig. S5 in Supporting Information II).

CONCLUSIONS

Deciphering the protein allostery has long been one of the grand challenges in molecular, structural, and computational biology. We elucidated allosteric hotspots and signaling pathways of the A_{2A}AR and other class A GPCRs by using the measure of betweenness centrality for each residue in protein structure network, the glycine scanning analysis, and the cross-correlation analysis based on the structural ensemble from MD simulations. Just like the role of native topology has been illuminated in the folding and unfolding mechanisms of proteins,[69–72] the success of analysis using graph representations of protein topology underscores the importance of native protein topology as one of the most critical determinants for intramolecular allosteric signaling. It is of special note that signals generated from protein dynamics, which include changes of inter-residue force, contact, or even local packing, are transmitted via the contacts formed between two neighboring residues. From the perspective of signal transduction, the betweenness centrality, defined with the number of parallel pathways on a given node, is physically a sensible way to quantify the amount of traffic on the node, thus to identify allosteric hotspots for a given protein structure. Given that residues of GPCRs associated with allostery and their signaling pathways are hard to capture using other conventional methods exploiting the information of sequence coevolution or variants of normal mode analysis (Figs. 6 and 7), the success of C_B-based analysis presented here is remarkable.

At the current stage, not only in the context of allosteric modulations in drug design[73–75] but also in the ligand binding (or release) induced conformational change in biological motors,[76,77] the importance of allostery in understanding the protein dynamics is highlighted more than ever. From the methodological perspective of this study, our C_B-based network analysis on protein structures is found quite powerful in identifying allosteric hotspots, and the results of analysis are in strong correlation with biochemical studies. The list of key residues for allostery and their cross-correlation identified here should be of great help to design experiments as well as contribute to our understanding to the dynamics of GPCRs. Our simple approach can not only be extended to study the allostery of other important proteins but also to study the allosteric communication within protein-protein or protein-RNA complexes.[43]

ACKNOWLEDGMENTS

The authors thank Korea Institute for Advanced Study (KIAS) and Korea Institute of Science and Technology Information (KISTI) Supercomputing Center for providing computing resources.

REFERENCES

1. Rosenbaum D, Rasmussen S, Kobilka B. The structure and function of G-protein-coupled receptors. Nature 2009;459:356–363.
2. Jensen A, Spalding T. Allosteric modulation of G-protein coupled receptors. Eur J Pharm Sci 2004;21:407–420.
3. Wess J. G-protein-coupled receptors: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. FASEB J 1997;11:346–354.
4. Kjelsberg M, Cotecchia S, Ostrowski J, Caron M, Lefkowitz R. Constitutive activation of the alpha 1B-adrenergic receptor by all amino acid substitutions at a single site. Evidence for a region which constrains receptor activation. J Biol Chem 1992;267:1430–1433.
5. Bond RA, Ijzerman AP. Recent developments in constitutive receptor activity and inverse agonism, and their potential for GPCR drug discovery. Trends Pharmacol Sci 2006;27:92–96.
6. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M, Eck MJ. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci USA 2008;105:2070–2075.
7. Yonath A. Antibiotics targeting ribosomes: Resistance, selectivity, synergism, and cellular regulation. Annu Rev Biochem 2005;74:649–679.
8. Galzi JL, Changeux JP. Neurotransmitter-gated ion channels as unconventional allosteric proteins. Curr Opin Struct Biol 1994;4:554–565.
9. Changeux JP. Allostery and the Monod-Wyman-Changeux model after 50 years. Annu Rev Biophys 2012;41:103–133.
10. Liu W, Chun E, Thompson A, Chubukov P, Xu F, Katritch V, Han G, Roth C, Heitman L, IJzerman A, Cherezov V, Stevens R. Structural basis for allosteric regulation of GPCRs by sodium ions. Science 2012;337:232–236.
11. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–299.
12. Süel G, Lockless S, Wall M, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Mol Biol 2002;10:59–69.
13. Dima RI, Thirumalai D. Determination of network of residues that regulate allostery in protein families using sequence analysis. Protein Sci 2006;15:258–268.
14. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks D, Sander C, Zecchina R, Onuchic J, Hwa T, Weigt M. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 2011;108:E1293–E1301.
15. Zheng W, Brooks BR, Doniach S, Thirumalai D. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure 2005;13:565–577.
16. Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci USA 2006;103:7664–7669.
17. Chennubhotla C, Bahar I. Markov propagation of allosteric effects in biomolecular systems: Application to GroEL-GroES. Mol Syst Biol 2006;2:msb4100075.
18. Balabin I, Yang W, Beratan D. Coarse-grained modeling of allosteric regulation in protein receptors. Proc Natl Acad Sci USA 2009;106:14253–14258.
19. Lee Y, Jeong LS, Choi S, Hyeon C. Link between allosteric signal transduction and functional dynamics in a multisubunit enzyme: *S*-adenosylhomocysteine hydrolase. J Am Chem Soc 2011;133:19807–19815.
20. Nygaard R, Frimurer T, Holst B, Rosenkilde M, Schwartz T. Ligand binding and micro-switches in 7TM receptor structures. Trends Pharmacol Sci 2009;30:249–259.
21. Freeman L. Centrality in social networks conceptual clarification. Soc Networks 1979;1:215–239.

22. Newman M. A measure of betweenness centrality based on random walks. Soc Networks 2005;27:39–54.
23. Jeong H, Mason SP, Barabasi A-L, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–42.
24. Del Sol A, Fujihashi H, Amoros D, Nussinov R. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. Mol Sys Biol 2006;2:2006.0019.
25. Edgar R. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004;5: 113.
26. Xu F, Wu H, Katritch V, Han G, Jacobson K, Gao Z, Cherezov V, Stevens R. Structure of an agonist-bound human A2A adenosine receptor. Science 2011;332:322.
27. Jaakola V, Griffith M, Hanson M, Cherezov V, Chien E, Lane J, IJzerman A, Stevens R. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science 2008;322:1211–1217.
28. Deupi X, Standfuss J. Structural insights into agonist-induced activation of G-protein-coupled receptors. Curr Opin Struct Biol 2011; 21:541–551.
29. Lebon G, Warne T, Edwards P, Bennett K, Langmead C, Leslie A, Tate C. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. Nature 2011;474:521–525.
30. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD, Jr. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. Biophys J 2006;90:L36–L38.
31. da Silveira C, Pires D, Minardi R, Ribeiro C, Veloso C, Lopes J, Meira W, Jr Neshich G, Ramos C, Habesch R, Marcelo M. Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. Proteins 2009;74:727–743.
32. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' netoworks. Nature 1998;393:440–442.
33. Albert R, Jeong H, Barabási A. Error and attack tolerance of complex networks. Nature 2000;406:378–382.
34. Jeong H, Tombor B, Albert R, Oltvai Z, Barabási A. The large-scale organization of metabolic networks. Nature 2000;407:651–654.
35. Balazsi G, Barabási A, Oltvai Z. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. Proc Natl Acad Sci USA 2005;102:7841–7846.
36. Vendruscolo M, Dokholyan N, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev E 2002;65:061910.
37. Atilgan A, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. Biophys J 2004;86:85–91.
38. Greene L, Higman V. Uncovering network systems within protein structures. J Mol Biol 2003;334:781–791.
39. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S. Network analysis of protein structures identifies functional residues. J Mol Biol 2004;344:1135–1146.
40. Bagler G, Sinha S. Assortative mixing in Protein Contact Networks and protein folding kinetics. Bioinformatics 2007;23:1760–1767.
41. Brinda K, Vishveshwara S. A network representation of protein structures: implications for protein stability. Biophys J 2005;89: 4159–4170.
42. Dokholyan N, Li L, Ding F, Shakhnovich E. Topological determinants of protein folding. Proc Natl Acad Sci USA 2002;99:8637.
43. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. Proc Natl Acad Sci USA 2009;106: 6620–6625.
44. Yao X-Q, Grant BJ. Domain-opening and dynamic coupling in the¡ i¿ α¡/i¿-subunit of heterotrimeric G proteins. Biophys J. 2013;105: L08–L10.
45. Miao Y, Nichols SE, Gasper PM, Metzger VT, McCammon JA. Activation and dynamic network of the M2 muscarinic receptor. Proc Natl Acad Sci USA 2013;110:10982–10987.
46. Dijkstra EW. A note on two problems in connection with graphs. Numerische Math 1959;1:269–271.
47. Brandes U. A faster algorithm for betweenness centrality. J Math Soc 2001;25:163–177.
48. Atilgan AR, Turgut D, Atilgan C. Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. Biophys J 2007;92:3052–3062.
49. Park K, Kim D. Modeling allosteric signal propagation using protein structure networks. BMC Bioinformatics 2011;12(Suppl 1):S23.
50. Daily MD, Gray JJ. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. PLoS Comput Biol 2009;5:e1000293.
51. Rasmussen S, DeVree B, Zou Y, Kruse A, Chung K, Kobilka T, Thian F, Chae P, Pardon E, Calinski D, Mathiesen J, Shah S, Lyons J, Caffrey M, Gellman S, Steyaert J, Skinoitis G, Weis W, Sunahara R, Kobilka B. Crystal structure of the β_2 adrenergic receptor-Gs protein complex. Nature 2011;477:549–555.
52. Katritch V, Cherezov V, Stevens RC. Structure-function of the G protein-coupled receptor superfamily. Annu Rev Pharmacol Toxicol 2013;53:531–556.
53. Hofmann K, Scheerer P, Hildebrand P, Choe H, Park J, Heck M, Ernst O. AG protein-coupled receptor at work: The rhodopsin model. Trends Biochem Sci 2009;34:540–552.
54. Suwa M, Sugihara M, Ono Y. Functional and structural overview of G-protein-coupled receptors comprehensively obtained from genome sequences. Pharmaceuticals 2011;4:652–664.
55. Ballesteros J, Weinstein H. Integrated methods for modeling G-protein coupled receptors. Methods Neurosci 1995;25:366–428.
56. Trzaskowski B, Latek D, Yuan S, Ghoshdastider U, Debinski A, Filipek S. Action of molecular switches in GPCRs-Theoretical and experimental studies. Curr Med Chem 2012;19(8):1090–1109.
57. Jacobson KA, Ohno M, Duong HT, Kim SK, Tchilibon S, Cesnek M, Holý A, Gao ZG. A neoceptor approach to unraveling microscopic interactions between the human A_{2A} adenosine receptor and its agonists. Chem Biol 2005;12:237–247.
58. Kim S, Gao Z, Van Rompaey P, Gross A, Chen A, Van Calenbergh S, Jacobson K. Modeling the adenosine receptors: comparison of the binding domains of A2A agonists and antagonists. J Med Chem 2003;46:4847–4859.
59. Horn F, Bettler E, Oliveira L, Campagne F, Cohen F, Vriend G. GPCRDB information system for G protein-coupled receptors. Nucleic Acids Res 2003;31:294–297.
60. Angel TE, Gupta S, Jastrzebska B, Palczewski K, Chance MR. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. Proc Natl Acad Sci USA 2009;106:14367–14372.
61. Halabi N, Rivoire O, Leibler S, Ranganathan R. Protein sectors: evolutionary units of three-dimensional structure. Cell 2009;138:774–786.
62. McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature 2012;138:138–142.
63. Shi L, Liapakis G, Xu R, Guarnieri F, Ballesteros J, Javitch J. β2 adrenergic receptor activation. J Biol Chem 2002;277:40989–40996.
64. Fritze O, Filipek S, Kuksa V, Palczewski K, Hofmann K, Ernst O. Role of the conserved NPxxY (x) 5, 6F motif in the rhodopsin ground state and during activation. Proc Natl Acad Sci USA 2003; 100:2290.
65. Katritch V, Cherezov V, Stevens R. Diversity and modularity of G protein-coupled receptor structures. Trends Pharmacol Sci 2011;33: 17–27.
66. Rosenkilde MM, Benned-Jensen T, Frimurer TM, Schwartz TW. The minor binding pocket: A major player in 7TM receptor activation. Trends Pharmacol Sci 2010;31:567–574.
67. del Sol A, Tsai C-J, Ma B, Nussinov R. The origin of allosteric functional modulation: multiple pre-existing pathways. Structure 2009; 17:1042–1050.
68. Cui Q, Karplus M. Allostery and cooperativity revisited. Protein Sci 2008;17:1295–1307.

69. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: The energy landscape perspective. Ann Rev Phys Chem 1997;48:539–600.
70. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci 1999;96:11305–11310.
71. Klimov DK, Thirumalai D. Native topology determines force-induced unfolding pathways in globular proteins. Proc Natl Acad Sci USA 2000;97:7254–7259.
72. Ferrara P, Caflisch A. Native topology or specific interactions: What is more important for protein folding? J Mol Biol 2001;306:837–850.
73. Swain J, Gierasch L. The changing landscape of protein allostery. Curr Opin Struct Biol 2006;16:102–108.
74. Goodey N, Benkovic S. Allosteric regulation and catalysis emerge via a common route. Nat Chem Biol 2008;4:474–482.
75. Hardy J, Wells J. Searching for new allosteric sites in enzymes. Curr Opin Struct Biol 2004;14:706–715.
76. Hyeon C, Lorimer GH, Thirumalai D. Dynamics of allosteric transition in GroEL. Proc Natl Acad Sci USA 2006;103:18939–18944.
77. Hyeon C, Onuchic JN. A structural perspective on the dynamics of Kinesin motors. Biophys J 2011;101:2749–2759.