BIOPHYSICS

Probing allosteric regulations with coevolution-driven molecular simulations

Francesco Colizzi[1]*[†] and Modesto Orozco[1,2]*

Protein-mediated allosteric regulations are essential in biology, but their quantitative characterization continues to posit formidable challenges for both experiments and computations. Here, we combine coevolutionary information, multiscale molecular simulations, and free-energy methods to interrogate and quantify the allosteric regulation of functional changes in protein complexes. We apply this approach to investigate the regulation of adenylyl cyclase (AC) by stimulatory and inhibitory G proteins—a prototypical allosteric system that has long escaped from in-depth molecular characterization. We reveal a surprisingly simple ON/OFF regulation of AC functional dynamics through multiple pathways of information transfer. The binding of G proteins reshapes the free-energy landscape of AC following the classical population-shift paradigm. The model agrees with structural and biochemical data and reveals previously unknown experimentally consistent intermediates. Our approach showcases a general strategy to explore uncharted functional space in complex biomolecular regulations.

INTRODUCTION

The free-energy landscape of proteins drives the functioning and regulation of cellular processes (*1–3*). Underlying these highly controlled activities is the balance among different conformational states, which is often regulated by allosteric effectors—ligands that produce a structural change in the target in a region distant from their binding site (*2*, *4*). Despite notable advances (*3*, *5–7*), the quantitative characterization and prediction of allosteric mechanisms continue to posit formidable challenges for both experiments and computations. The molecular underpinnings of allosteric changes in protein structures are often hindered by averaged metrics, the transient nature of the species involved (*8*), the difficulty to reproduce biological events in vitro (*9*), or the great deal of computational power required to model uncharted free-energy surfaces (*9*, *10*). Nonetheless, allosteric regulations are widespread in biology (*11–18*), and the characterization of the underlying free energy is critical to understand and control the functional conformational landscape of biomolecules (*19*). The objective of this work is thus to devise a computational procedure to efficiently explore the functional free-energy landscape of a protein with and without the allosteric effector bound, thereby gaining insight into how conformational equilibria translates into function. We focus on the conserved regulation of adenylyl cyclase (AC) by stimulatory and inhibitory G proteins—a prototypical example of protein-protein allostery at the center of the G protein–coupled receptor (GPCR) signaling cascade, the target of nearly half of all drugs. Despite extensive molecular research conducted on this system (*20–22*), there are still open questions.

When hormones or drugs bind to GPCRs, they activate G proteins, which, in turn, activate or inhibit AC (*20–22*). ACs control the rate of conversion of adenosine 5′-triphosphate (ATP) into 3′,5′-cyclic adenosine monophosphate (cAMP)—the second messenger that, by interacting with protein kinase A, triggers the phosphorylation of a myriad of downstream targets (*23*). The most conserved regions of ACs are the pseudo-symmetric catalytic domains (called C_1 and C_2) that work as obligate dimers with the active site at the dimer interface, a feature rarely found in other enzymes (*24*). This unique feature makes the catalytic rate markedly affected by structural movements at the seam of the two domains (*22*, *24*). X-ray crystallography (*25–27*) and cryo–electron microscopy (cryo-EM) (*22*, *28*) experiments have detailed two conformational states in AC structure. They are coupled to the formation of the complex with stimulatory Gα_s protein and include the reorientation of AC catalytic domains, together with a small displacement of an α helix opposite to the binding site of stimulatory Gα_s (*22*, *24–27*).

The high versatility of signal transduction encoded in AC structure (*22*, *24*, *29*) suggests, however, that the known states are only snapshots of a much wider and diverse set of functional ensembles that still have to be captured and quantified. In this context, computer simulations (*30*) could bring about a major productivity leap. Recently, insight has been generated from long molecular dynamics (MD) simulations, confirming the high structural plasticity of AC and providing a general blueprint of AC dynamics in a variety of biological contexts (*31–37*). Although extremely informative, these studies have offered only a qualitative description of the complicated link between effector binding and AC dynamics, raising challenging questions on the mechanisms and the free-energy landscape governing the functional ensemble of AC structures.

To face these challenges, we devised a multiscale MD approach guided by coevolutionary data (*38*) that allowed the free-energy landscape of a fully solvated atomistic model of AC with and without the bound G protein to be reconstructed. Coevolutionary-based approaches have been used to predict native structures in proteins (*39*, *40*) and nucleic acids (*41*, *42*) and to generate conformational ensembles (*43–45*). Our work is unique in that it leverages coevolutionary information to reduce the complexity of the configurational space in protein-protein regulations, thereby focusing only on the functional free-energy landscape obtained from residue coevolution. By doing so, we expand and manipulate AC structural

[1]Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute of Science and Technology (BIST), Carrer de Baldiri Reixac 10, Barcelona 08028, Spain. [2]Departament de Bioquímica i Biomedicina, Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal 647, Barcelona 08028, Spain.
*Corresponding author. Email: francesco.colizzi@irbbarcelona.org (F.C.); modesto.orozco@irbbarcelona.org (M.O.)
†Present address: Institute of Marine Sciences, ICM-CSIC, Passeig Marítim de la Barceloneta 37-49, Barcelona 08003, Spain.

ensembles consistently with—and beyond—experiments. We focus on human AC type 5 (hAC5) as quintessential allosteric unit controlled by both stimulatory and inhibitory Gα proteins (20). We find AC populating two main conformational ensembles with all the existing experimental structures falling into just one of these ensembles. Notably, AC shifts from one ensemble to the other depending on which G protein it binds to. The results complement, and are in agreement with, structural (22) and biochemical (29, 46) data, and extend our understanding of the mechanisms of protein-induced allostery in AC. Furthermore, they provide a general framework for investigating and manipulating complex biomolecular regulations.

RESULTS AND DISCUSSION
Coevolutionary analysis captures functional domain reorientations
We performed direct coupling analysis (DCA; see Methods) (40, 47) on a multiple sequence alignment of hAC5 homologs, selecting those coevolving pairs that are not in contact in the native structure, and which thus bear information on possible alternative, unknown, conformations (i.e., nontrivial pairs; Fig. 1A, red dots) (43). Nontrivial pairs appear between C_1 and C_2 catalytic domains (Fig. 1A, left), suggesting that interdomain movements are linked to function.
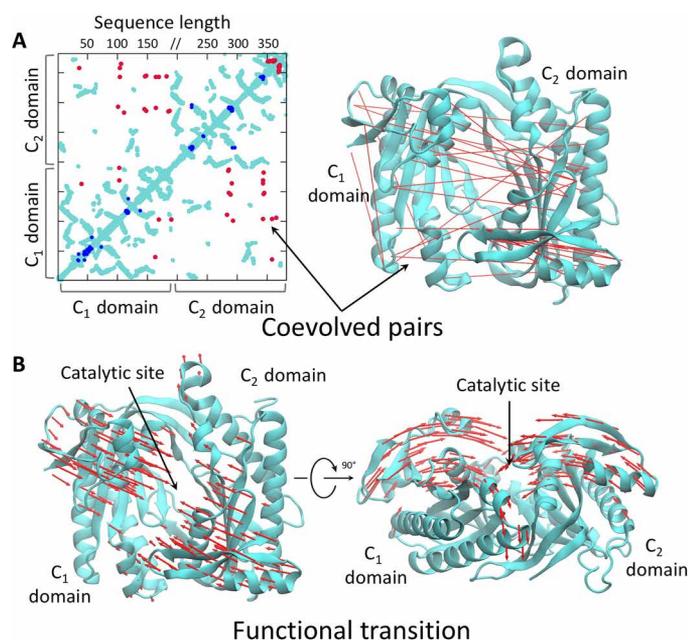

Fig. 1. Residue coevolution is used to explore the functional conformational landscape of hAC5. (**A**) Contact map (left) of hAC5 showing residues that are within 8 Å in the native structure (cyan dots) together with coevolved pairs of residues that are not in contact in the native structure (red dots) and that might thus be informative of alternative protein conformations. Coevolved pairs of residues that are in contact in the native structure are shown as blue dots. Right: Structural representation (red lines) of nontrivial coevolved pairs in hAC5 (cyan ribbons). For the sake of clarity, only the top 30 DCA contacts for hAC5 are shown. The quantitative threshold used to determine which contacts were included in the coarse-grained simulations is described in Methods. (**B**) Deformation arrows (in red) showing the motion corresponding to the first principal component of the functional transition reconstructed from DCA and coarse-grained simulations. The closure movement of C_1 and C_2 domain modulates the shape and accessibility of the catalytic site.

DCA scores were filtered and introduced as ensemble restraints in coarse-grained discrete MD (dMD) simulations (43, 48–51), which allowed us not only to detect functionally relevant conformations but also to generate trajectories connecting them (43). The trajectories resulted in hAC5 conformational transitions dominated by the opening and closing of the catalytic site as shown by the analysis of its principal components (Fig. 1B). The closure movement of C_1 and C_2 domain modulates the shape and accessibility of the catalytic site to ATP, providing a molecular glimpse on the major regulation mechanism of AC activity (22, 24–27).

Functional free-energy landscape from path metadynamics
To probe the energetic feasibility of the coevolution-based dMD transition, we back mapped the coarse-grained transition into a fully atomistic one, characterizing its free-energy landscape with path-based (52) metadynamics (53, 54) simulations in explicit solvent (see Methods and the Supplementary Materials). We observed that the apo (ligand-free) hAC5 populated two free-energy basins with a similar well-depth of 4 kcal/mol (Fig. 2 and fig. S3). Such free-energy minima corresponded to a sparse "open" ensemble (Fig. 2, A and B) and to a well-defined closed conformation of the catalytic domain and active site. Notably, the open ensemble included the experimental structures of activated AC bound to $Gα_s$ together with an ATP analog (P-site inhibitor) and the activator forskolin (1cjk) (26), a forskolin derivative alone (1azs) (25), and a guanosine 5′-triphosphate (GTP)–based substrate analog and forskolin (6r4o) from cryo-EM experiments (28) of the full-length membrane AC (Fig. 2A, gray circles). As the above AC structures were experimentally solved in complex with the stimulatory $Gα_s$, we label this open ensemble as the "active conformation"; the residues forming the active site are apart from each other and appear in an ATP-binding competent state. Although ATP was not present in our simulations, note that an ATP-bound conformation is a free-energy minimum encoded in AC intrinsic dynamics.

Moving along the functional transition, we observed a metastable state (labeled 2 in Fig. 2) with reduced interdomain distance than the open ensemble. This intermediate conformation features a salt bridge between D518 and R1208 from C_1 and C_2 domain, respectively. D518 has been experimentally observed interacting with catalytic Mg^{2+} and R1208 with the α-phosphate of ATP analogs and is thus essential for ATP binding and catalysis (26). The desolvation of D518 and R1208 side chains likely contributes to the energy barrier for reaching this intermediate state from the open ensemble. Further progression along the functional reaction coordinate leads the system to a compacted closed conformation (labeled 3 in Fig. 2) with the interdomain distance dropping below 25 Å. Such a closed conformation is characterized by an additional bidentate salt bridge between D474 and E596 and the catalytic K1244. This interaction "seals" AC into a conformation with no accessible binding site for ATP (Fig. 2, A and B). As ATP binding is essential for cAMP production, this conformation is catalytically inactive. We remark that closed conformations of AC have not yet been observed experimentally, likely because the structure of AC could only be solved either in complex with the stimulatory $Gα_s$ (25, 26, 28, 55–57) or with the active site occupied by the activator forskolin (27). These results motivate and challenge the design of future experimental research. Overall, path metadynamics allowed the atomistic description of functional transitions in hAC5 and the efficient reconstruction of the underlying free-energy landscape with an estimated error in the order of k_BT (Fig. 2C).
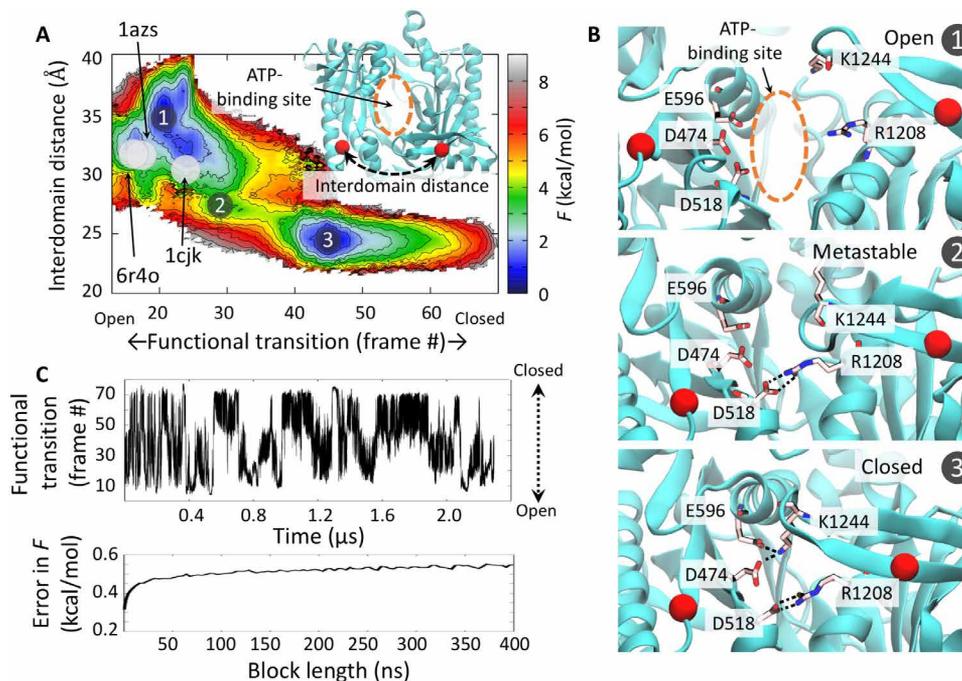
Fig. 2. Mapping the coevolution-driven conformational transition of hAC5 with fully atomistic metadynamics simulations. (**A**) Free-energy landscape of hAC5 conformational change reconstructed as a function of the progress along the functional transition and the distance between the C_1 and C_2 domain, using Asn[558] and Met[1250] as reference (red spheres in the structure inset). The encircled numbers correspond to the open, metastable, and closed hAC5 states shown in panel. Gray transparent circles with corresponding PDB codes represent the position of known experimental AC structures in the explored conformational space. In the inset, the structure of hAC5 is represented by cyan ribbons, with ATP-binding site and interdomain distance highlighted. (**B**) Snapshots of the metadynamics trajectory showing key residues of the catalytic site (carbon atoms in pink) and the main interactions (dotted black lines) formed at the interface between C_1 and C_2 domain in an open (1), metastable (2), and closed conformation (3). The residues highlighting the interdomain distance (red spheres) are shown as a reference; hAC5 is represented by cyan ribbons. (**C**) Diffusion of the system in the functional-transition space as a function of simulated time (top), highlighting the exploration of the same region multiple times. Error in the reconstructed free energy with different length of block averages (bottom).

Population shifts upon G protein binding

To assess the effect of stimulatory and inhibitory G proteins on the regulation of hAC5 structural plasticity, we perturbed the free-energy landscape of hAC5 with the presence of either stimulatory $G\alpha_s$ or inhibitory $G\alpha_i$ (Fig. 3). The stimulatory G protein, $G\alpha_s$, binds within a cleft in the C_2 domain of AC, while $G\alpha_i$ binds within the opposite cleft in the C_1 domain (Fig. 3) (*29*). When $G\alpha_s$ binds to hAC5, the conformational ensemble of AC is shifted to the open conformation that becomes ~6 kcal/mol more stable than the closed state (Fig. 3A, blue lines, and figs. S1 and S3). Notably, this open conformation belongs to the same ensemble comprising the experimental structures of AC complexed with stimulatory $G\alpha_s$, thus validating the theoretical framework used here (see Fig. 2A for reference). In notable contrast, however, when we simulated hAC5 bound to the inhibitory $G\alpha_i$, the open/closed equilibrium shifted toward the closed conformation of AC that becomes ~8 kcal/mol more stable than the open state (Fig. 3A, black lines, and figs. S2 and S3). The closure of the ATP-binding site is consistent with biochemical data indicating that P-site inhibitors (ATP analogs) bind with greatly reduced affinity to AC in the presence of $G\alpha_i$ (*46*). This behavior suggests a straightforward mechanism of AC regulation by G proteins. That is, when $G\alpha_s$ binds to hAC5 (Fig. 3B), the open ensemble of conformations is selected and the cyclase becomes competent to host ATP in its active site; vice versa, when $G\alpha_i$ binds to hAC5 (Fig. 3C, the binding of $G\alpha_i$ is symmetrically opposed to $G\alpha_s$), the

ensemble of hAC5 conformations shifts toward a closed state that impedes or perturbs ATP binding, thus resulting in the inhibition of cAMP production.

Multiple pathways of information transfer

The structural dynamics of ACs systems was further investigated with community network analysis (see Methods) (*58*) to identify and compare major pathways of signal transduction and allosteric communication (Fig. 4). Analyzing the community network of the hAC5/$G\alpha_s$ system, a major communication path is found between the binding site of $G\alpha_s$ (community #1) and the community (#6) including the small α4 and α7 helices, which line the extension of the catalytic site in the C_1 domain (Fig. 4A). The communication is mediated by correlated motions in the β sheet (β1, β4, and β5) connecting the two communities, as found in other globular proteins (*59*), and in the hAC5/$G\alpha_s$ system, they further mediate a weaker allosteric communication between communities 3 and 4, also lining the catalytic site. The dynamical network of the hAC5/$G\alpha_i$ system shows quite different features, as the communication from the binding site of inhibitory $G\alpha_i$ (community #3) propagates to the rest of the protein through the α-helix 3 of C_1 domain (Fig. 4B). Particularly, $G\alpha_i$ allosterically modulates the binding site of $G\alpha_s$ (community #1), suggesting a mechanism to inhibit cAMP production by interfering with binding of the stimulatory $G\alpha_s$, as it has been proposed previously (*29*, *32*, *35*). Therefore, the binding of $G\alpha_i$ not
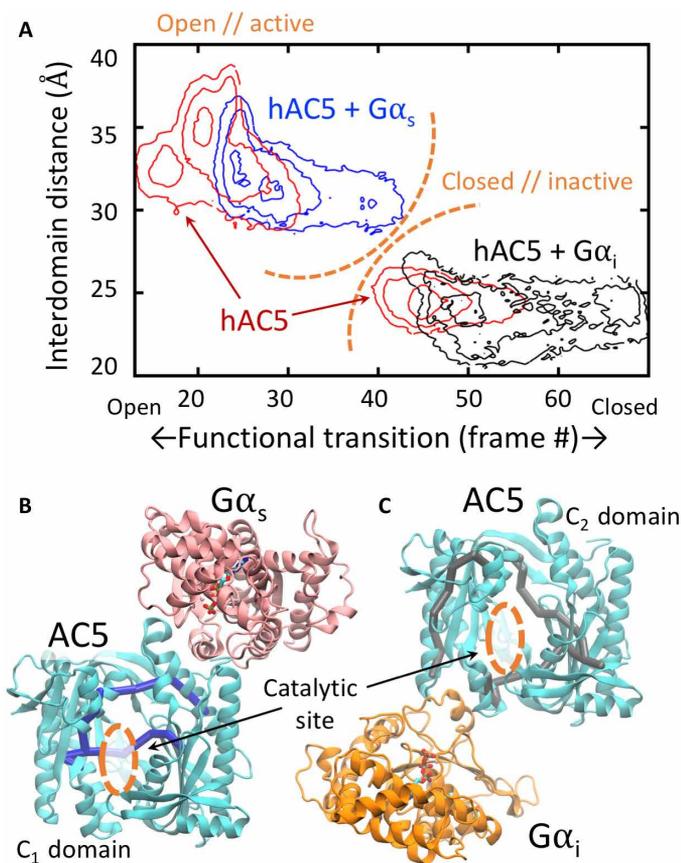
Fig. 3. Modulation of hAC5 functional transition by the binding of stimulatory and inhibitory G proteins. (**A**) Comparison of the free-energy minima populated by hAC5 with no regulatory protein bound (red plot), in complex with stimulatory Gα$_s$ (blue plot), and in complex with inhibitory Gα$_i$ (black plot); contours drawn at 1 kcal/mol intervals from 0 to 3 kcal/mol, after alignment of each minimum to zero. hAC5 alone can populate both closed and open states, the binding of Gα$_s$ stabilizes an open state, while the binding of Gα$_i$ favors the closed conformation of hAC5. (**B**) Structural representation of the complex between hAC5 (cyan ribbons) and Gα$_s$ (pink ribbons); blue tubes on hAC5 represent main pathways of signal transduction between C$_1$ and C$_2$ domain from community network analysis (see text). (**C**) Structural representation of the complex between hAC5 (cyan ribbons) and Gα$_i$ (orange ribbons); gray tubes on hAC5 represent main pathways of signal transduction between C$_1$ and C$_2$ domains from community network analysis (see text).

Fig. 4. Pathways of signal transduction and allosteric communication in hAC5 from community network analysis. Community network representation for hAC5 bound to (**A**) Gα$_s$ and (**B**) Gα$_i$. The communities are shown in different colors as filled circles (left) or cartoon structures (right). Major communication pathways in the network are identified by the intercommunity connections shown as lines, with width proportional to the cumulative betweenness of intercommunity edges (left). Residues forming salt bridges in the metastable and closed state are shown under the belonging community (left). The binding site of G proteins is highlighted by dashed gray lines. Binding of Gα$_s$ to the C$_2$ domain is communicated to the C$_1$ domain via the β-sheet motif formed by β1, β4, and β5 (intercommunity connections 1-6 and 3-4). The modulation of hAC5 dynamics by Gα$_i$ is achieved through a larger network reaching the C$_1$ domain mainly passing through the helix α3 (intercommunity connection 3-2). As highlighted by the community network, the binding of one Gα subunit can produce allosteric effects at the binding site for the other, with the binding of Gα$_i$ having the major allosteric modulation on the dynamics of Gα$_s$ binding site.

only selects the closed state of hAC5 catalytic domain but also triggers information transfer to the binding site of Gα$_s$, likely modulating its capability to activate the cyclase. Furthermore, the residues participating in the salt bridges observed in the intermediate and closed state are involved in different communication pathways depending on which G protein AC binds to. In the hAC5/Gα$_s$ complex, residues D474 and E596 are the critical nodes connecting communities 6 and 4, respectively, while D518 is the critical node connecting communities 4 to 3. Residues K1244 and R1208 belong to community 5 (Fig. 4A, left). There is no direct communication between communities 5 and 6, 4, likely reflecting the observation that the free-energy landscape of the hAC5/Gα$_s$ complex is shifted toward the open state where no salt bridges are formed. In the hAC5/Gα$_i$ complex, residues K1244 and R1208 fall into communities 5 and 4, respectively (Fig. 4B, left). In this complex, we observe a direct communication between communities 3 and 4 that could
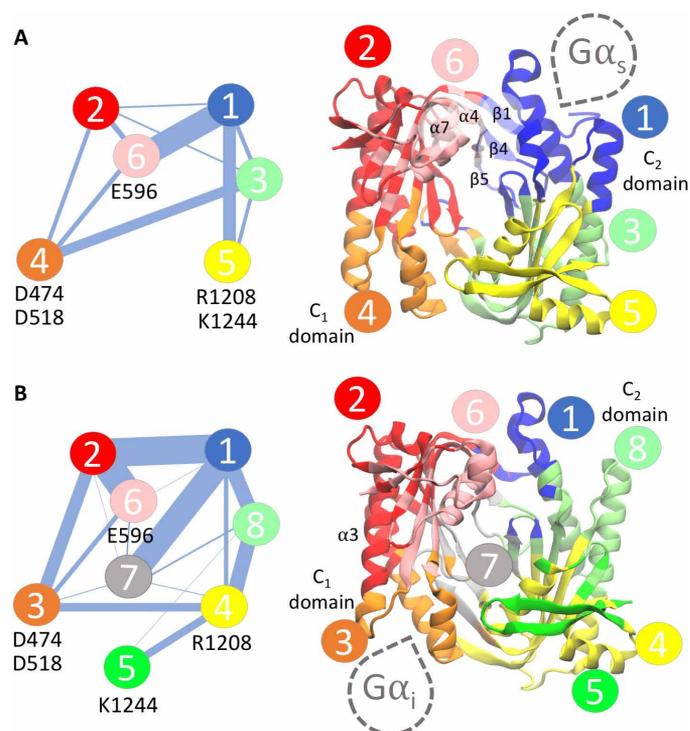
mediate the formation of the salt bridge between D518 and R1208 featured in the intermediate state (Fig. 2B). Notably, community 5, including the catalytic K1244, appears to communicate mostly with community 4, suggesting a possible flow of information between intermediate and closed state. We further note that the asymmetry in the communication pathways activated when Gα$_s$ binds AC compared to Gα$_i$ binding could be related to the asymmetry in the relative (de)stabilization of open/closed states by the two G proteins (with Gα$_s$ stabilizing by ~6 kcal/mol the open conformation, while Gα$_i$ stabilizing by ~8 kcal/mol the closed one). In particular, a qualitative comparison of the networks shown in Fig. 4 (A and B) suggests that the binding of Gα$_i$ to AC triggers a much wider flow of information with high correlation (or energy of interaction) (58) between nodes than the binding of Gα$_s$ to AC. Last, the large number of highly conserved residues at the interface between communities (critical nodes) suggests that the response to external stimuli (e.g., to inhibitory Gα$_i$) among different AC isoforms can be achieved by

sequence changes as well as by specific binding to effectors that, in turn, modulate intracommunity dynamics (60).

By combining recent breakthroughs in coevolutionary analysis with multiscale modeling and free-energy methods, we have dissected the allosteric regulation of AC intrinsic dynamics by G proteins—a fundamental process in biology and pharmacology. Our model reveals quantitative details on the regulation mechanism and demonstrates that signal transduction in AC/G protein systems operates through the selective (de)stabilization of the particular state to which ATP preferentially binds following the classical "Monod-Wyman-Changeux" population-shift paradigm (2, 61). These results create a common reading frame among multiple lines of experimental data and provide an unprecedented spatiotemporal resolution on the molecular mechanisms regulating cAMP generation. A possible limitation of the present work is that coevolutionary information was encoded only in the dynamics of AC, thus assuming that the activated GTP-bound state of G proteins fluctuates around the observed crystallographic structure (22).

Overall, the general procedure outlined here can be used to probe other allosteric regulations in uncharted conformational space for a wide range of complex systems. Advances in genomic sequencing make the procedure applicable to thousands (43) of macromolecules for which functionally relevant transitions can be efficiently perturbed. Molecular design strategies can be combined with the approach presented here to interrogate desirable functional intermediates, change the kinetics, or reprogram the mechanism of these systems with far-reaching consequences for pharmaceutical and biotechnological applications.

METHODS
3D structural models
The catalytic domains of hAC5 (UniProt: O95622) were built by homology modeling using SWISS-MODEL (62), based on up to 98% sequence identity with the crystal structure of the mammalian hybrid AC5/AC2 in complex with $G\alpha_s$ [Protein Data Bank (PDB) ID: 1AZS]; GTP-activated human $G\alpha_s$ (UniProt: P63092) in complex with hAC5 was built from the same PDB template. The crystal structure of GTP-activated human $G\alpha_i$ (UniProt: P63096; 2GTP) was used to generate the complex with hAC5 that resulted in $G\alpha_i$ binding the C_1 domain in an orientation symmetrically opposed to $G\alpha_s$ with respect to AC5 in the 1AZS complex, as suggested by mutagenesis data (29). Myristoylation of $G\alpha_i$ was modeled following the procedure described in (63).

Coevolutionary analysis and coarse-grained simulations
The pairwise coevolved amino acid positions correlate strongly with spatial proximity in the three-dimensional (3D) space (38–40, 47) and can be used to fuel the search for biologically relevant conformational ensembles (43–45) and to identify functionally relevant transitions in proteins (43, 44) and nucleic acids (41, 42). Multiple sequence alignment (1261 hits) from the clustered UniProt database (uniclust30_2018_08) (64) was generated using hAC5 (UniProt code: O95622) as query with HHblits (65), and DCA (47) was used to measure residue-residue coevolution with default parameters. DCA outputs a direct information (DI) score per pair of residues that was then filtered and used as input for the coarse-grained dMD simulations as described by Sfriso et al. (43). Briefly, given a DI-ranked list of coevolution pairs, we keep for further analysis only the first n pairs

(n = 1000 in the case of AC) that maximize the Matthew's correlation coefficient resulting from the prediction of contacts (<10 Å) in the initial structure. Intuitively, in this step, we extend to the number of DCA contacts that are still informative about the initial structure. For the n pair selected, we test the accessibility of each residue pair in the structure by means of dMD pulling simulations where DCA pairs are brought close in space (one independent dMD simulation for each pair). By doing so, we filter the DCA output to remove uninformative or impossible contact pairs; individual trajectories are accepted only if they show better coincidence with coevolution information than a given threshold. For this, we check whether coevolutionary contacts are spontaneously established along the pulling trajectory and compute receiver operating characteristic (ROC) curves to quantify the agreement between the conformations generated and the list of n coevolution pairs. The area under the resulting ROC curve (AUC) provides a means to compare and rank the coherence between trajectories and the coevolutionary fingerprint. We retain instances exceeding 1.5 of the interquartile range in the AUC distribution (16 generated models in the case of AC), and the corresponding pairs of residues are incorporated as weighed (43) square potentials (904 wells were added to the native 9631 wells in the case of AC) into a multiple structure-based model (SBM). Coevolution pairs are thus reflected in the multiple SBM by favorable energy interactions that guide an ad hoc sampling strategy (43) in the exploration of the conformational landscape with dMD. Implicitly, this approach filters noise in the DCA signal and reveals the protein ensemble encoded by coevolution. The dMD simulations are clustered and analyzed to provide a trajectory that is representative of the functional conformational landscape. Last, the trajectory was discretized in 80 equidistant frames (66), which were sufficient to describe the functional transition with the necessary resolution for metadynamics. We note that our approach is similar to the one proposed by Morcos et al. (44), with the main difference being the filtering step of the DCA contacts. While Morcos et al. (44) directly incorporate top-ranked DCA pairs as energy minima of the SBM, our protocol includes the filtering of coevolution pairs with the dMD pulling simulations mentioned above. We have shown (43) that filtering of DCA contacts is not critical when abundant sequences (>10,000) are available, yielding strong evolutionary signal. However, when fewer homologs are aligned (in the order of few thousands), conformational transitions could not be modeled without filtering DCA contacts due to noise in the coevolution map. In our experience, few high-quality coevolved pairs are thus necessary to robustly guide protein dynamics, making the detection of these constraints decisive.

System setup and MD simulations
Each system was solvated with a 1-nm-thick truncated octahedron box of TIP3P (67) water molecules with periodic boundary conditions. Equations of motion were integrated with a time step of 2 fs. For all nonbonded interactions, the real space cutoff was set to 1.0 nm, and the electrostatic long-range interactions were treated using the default particle-mesh Ewald settings (68). Bonds involving hydrogen atoms were constrained using the LINCS algorithm (69). After minimization and thermalization in the canonical ensemble (NVT), each system was further equilibrated at constant pressure and temperature (1 atm, 298 K) for 100 ns; metadynamics production runs were then generated in the isothermal–isobaric ensemble (NPT) using the stochastic velocity rescaling thermostat (70) and the

Parrinello-Rahman barostat (*71*). The active site of Gα proteins was modeled with one crystallographic Mg^{2+} ion and a GTP molecule. N-terminal myristoylation was present in Gα$_i$ and modeled following the procedure described in (*63*). AMBER parameters for Mg^{2+} (*72*), GTP (*73*), and myristoyl group (*63*) were used. All simulations were run using GROMACS 2018.3 (*74*) and PLUMED 2.4.3 (*75*) with the Amber ff99SB-ILDN force field (*76*).

Atomistic metadynamics simulations

Well-tempered metadynamics (*77*) was used to reconstruct the underlying free energy as a function of two collective variables (CVs) or reaction coordinates based on the functional path generated from the coevolutionary-driven coarse-grained simulations. In this framework, the microscopic coordinates of the system, *q*, are mapped in the CV space by *s*(*q*), which measures the progress along the functional path (*52*), and *z*(*q*), which measures the distance from the preassigned path. Using these variables, one can explore the free-energy landscape between an initial and final state and can find low free-energy pathways connecting them—pathways that, in turn, can be different from the originally assigned one (*52*). Proper reweighting (*78*) allowed the resulting free-energy landscape to be projected on the interdomain distance and *s*(*q*) (see also figs. S1 to S3). We ran 2.3 μs of well-tempered metadynamics simulations for each system.

Community network analysis

Allosteric network in the AC systems was examined through community network analysis (*58*) and visualized (*79*) with VMD (*80*). Briefly, the metadynamics trajectory of the AC systems is used to group Cα carbons (nodes) into communities—a set of nodes that move in concert with each other. Edges between pair of nodes are drawn if the average internode distance is below a certain threshold (5 Å); edge distances between node *i* and *j* have correlation-based weights, $w_{ij} = -\log(|C_{ij}|)$, which define the probability of information transfer across a given edge. Major communication pathways are identified by the edge betweenness, the number of shortest paths that cross a given edge (*58*).

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abj0786

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. M. Karplus, J. Kuriyan, Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6679–6685 (2005).
2. J. P. Changeux, A. Christopoulos, Allosteric modulation as a unifying mechanism for receptor function and regulation. *Cell* **166**, 1084–1102 (2016).
3. T. Xie, T. Saleh, P. Rossi, C. G. Kalodimos, Conformational states dynamically populated by a kinase determine its function. *Science* **370**, eabc2754 (2020).
4. N. M. Goodey, S. J. Benkovic, Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* **4**, 474–482 (2008).
5. R. A. Langan, S. E. Boyken, A. H. Ng, J. A. Samson, G. Dods, A. M. Westbrook, T. H. Nguyen, M. J. Lajoie, Z. Chen, S. Berger, V. K. Mulligan, J. E. Dueber, W. R. P. Novak, H. El-Samad, D. Baker, De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019).
6. K. Y. M. Chen, D. Keri, P. Barth, Computational design of G protein-coupled receptor allosteric signal transductions. *Nat. Chem. Biol.* **16**, 77–86 (2020).
7. G. Mattedi, S. Acosta-Gutiérrez, T. Clark, F. L. Gervasio, A combined activation mechanism for the glucagon receptor. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15414–15422 (2020).
8. M. Gruebele, Protein dynamics in simulation and experiment. *J. Am. Chem. Soc.* **136**, 16695–16697 (2014).
9. S. Bottaro, K. Lindorff-Larsen, Biophysical experiments and biomolecular simulations: A perfect match? *Science* **361**, 355–360 (2018).
10. R. D. Malmstrom, A. P. Kornev, S. S. Taylor, R. E. Amaro, Allostery through the computational microscope: cAMP activation of a canonical signalling domain. *Nat. Commun.* **6**, 7588 (2015).
11. D. Hilger, M. Masureel, B. K. Kobilka, Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).
12. N. R. Latorraca, A. J. Venkatakrishnan, R. O. Dror, GPCR dynamics: Structures in motion. *Chem. Rev.* **117**, 139–155 (2017).
13. D. D. Boehr, R. Nussinov, P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
14. A. H. Phillips, A. J. Schoeffler, T. Matsui, T. M. Weiss, J. W. Blankenship, K. Zobel, A. M. Giannetti, E. C. Dueber, W. J. Fairbrother, Internal motions prime cIAP1 for rapid activation. *Nat. Struct. Mol. Biol.* **21**, 1068–1074 (2014).
15. L. Cupellini, D. Calvani, D. Jacquemin, B. Mennucci, Charge transfer from the carotenoid can quench chlorophyll excitation in antenna complexes of plants. *Nat. Commun.* **11**, 662 (2020).
16. M. Dutta, S. P. Gilbert, J. N. Onuchic, B. Jana, Mechanistic basis of propofol-induced disruption of kinesin processivity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023659118 (2021).
17. P. H. Von Hippel, BIOCHEMISTRY: Completing the view of transcriptional regulation. *Science* **305**, 350–352 (2004).
18. C. G. Kalodimos, N. Biris, A. M. J. J. Bonvin, M. M. Levandoski, M. Guennuegues, R. Boelens, R. Kaptein, Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386–389 (2004).
19. N. V. Dokholyan, Controlling allosteric networks in proteins. *Chem. Rev.* **116**, 6463–6487 (2016).
20. R. K. Sunahara, C. W. Dessauer, A. G. Gilman, Complexity and diversity of mammalian adenylyl cyclases. *Annu. Rev. Pharmacol. Toxicol.* **36**, 461–480 (1996).
21. R. K. Sunahara, R. Taussig, Isoforms of mammalian adenylyl cyclase: Multiplicities of signaling. *Mol. Interv.* **2**, 168–184 (2002).
22. B. Khannpnavar, V. Mehta, C. Qi, V. Korkhov, Structure and function of adenylyl cyclases, key enzymes in cellular signaling. *Curr. Opin. Struct. Biol.* **63**, 34–41 (2020).
23. J. M. Gancedo, Biological roles of cAMP: Variations on a theme in the different kingdoms of life. *Biol. Rev.* **88**, 645–668 (2013).
24. J. U. Linder, J. E. Schultz, Versatility of signal transduction encoded in dimeric adenylyl cyclases. *Curr. Opin. Struct. Biol.* **18**, 667–672 (2008).
25. J. J. G. Tesmer, R. K. Sunahara, A. G. Gilman, S. R. Sprang, Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gs·GTPS. *Science* **278**, 1907–1916 (1997).
26. J. J. G. Tesmer, R. K. Sunahara, R. A. Johnson, G. Gosselin, A. G. Gilman, S. R. Sprang, Two-metal-ion catalysis in adenylyl cyclase. *Science* **285**, 756–760 (1999).
27. G. Zhang, Y. Liu, A. E. Ruoho, J. H. Hurley, Structure of the adenylyl cyclase catalytic core. *Nature* **386**, 247–253 (1997).
28. C. Qi, S. Sorrentino, O. Medalia, V. M. Korkhov, The structure of a membrane adenylyl cyclase bound to an activated stimulatory G protein. *Science* **364**, 389–394 (2019).
29. C. W. Dessauer, J. J. G. Tesmer, S. R. Sprang, A. G. Gilman, Identification of a Giα binding site on type V adenylyl cyclase. *J. Biol. Chem.* **273**, 25831–25839 (1998).
30. M. Karplus, J. A. McCammon, Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).
31. E. Frezza, J. Martin, R. Lavery, A molecular dynamics study of adenylyl cyclase: The impact of ATP and G-protein binding. *PLOS ONE* **13**, e0196207 (2018).
32. S. C. van Keulen, U. Rothlisberger, Exploring the inhibition mechanism of adenylyl cyclase type 5 by n-terminal myristoylated Gα$_{i1}$. *PLOS Comput. Biol.* **13**, e1005673 (2017).
33. S. C. Van Keulen, D. Narzi, U. Rothlisberger, Association of both inhibitory and stimulatory gα subunits implies adenylyl cyclase 5 deactivation. *Biochemistry* **58**, 4317–4324 (2019).
34. N. J. Bruce, D. Narzi, D. Trpevski, S. C. van Keulen, A. G. Nair, U. Röthlisberger, R. C. Wade, P. Carloni, J. H. Kotaleski, Regulation of adenylyl cyclase 5 in striatal neurons confers the ability to detect coincident neuromodulatory signals. *PLOS Comput. Biol.* **15**, e1007382 (2019).
35. E. Frezza, T. M. Amans, J. Martin, Allosteric inhibition of adenylyl cyclase type 5 by g-protein: A molecular dynamics study. *Biomolecules* **10**, 1330 (2020).
36. K. Amunts, C. Ebell, J. Muller, M. Telefont, A. Knoll, T. Lippert, The Human Brain Project: Creating a European research infrastructure to decode the human brain. *Neuron* **92**, 574–581 (2016).
37. D. Narzi, S. C. van Keulen, U. Röthlisberger, Gα$_{i1}$ inhibition mechanism of ATP-bound adenylyl cyclase type 5. *PLOS One* **16**, e0245197 (2021), 10.1371/journal.pone.0245197.
38. D. De Juan, F. Pazos, A. Valencia, Emerging methods in protein co-evolution. *Nat. Rev. Genet.* **14**, 249–261 (2013).
39. J. I. Sułkowska, F. Morcos, M. Weigt, T. Hwa, J. N. Onuchic, Genomics-aided structure prediction. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10340–10345 (2012).
40. F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293–E1301 (2011).

41. F. Cuturello, G. Tiana, G. Bussi, Assessing the accuracy of direct-coupling analysis for RNA contact prediction. *RNA* **26**, 637–647 (2020).

42. C. Weinreb, A. J. Riesselman, J. B. Ingraham, T. Gross, C. Sander, D. S. Marks, 3D RNA and functional interactions from evolutionary couplings. *Cell* **165**, 963–975 (2016).

43. P. Sfriso, M. Duran-Frigola, R. Mosca, A. Emperador, P. Aloy, M. Orozco, Residues coevolution guides the systematic identification of alternative functional conformations in proteins. *Structure* **24**, 116–126 (2016).

44. F. Morcos, B. Jana, T. Hwa, J. N. Onuchic, Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20533–20538 (2013).

45. L. Sutto, S. Marsili, A. Valencia, F. L. Gervasio, From residue coevolution to protein conformational ensembles and functional dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13567–13572 (2015).

46. C. W. Dessauer, M. Chen-Goodspeed, J. Chen, Mechanism of $G\alpha_i$-mediated inhibition of type V adenylyl cyclase. *J. Biol. Chem.* **277**, 28823–28829 (2002).

47. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 67–72 (2009).

48. E. A. Proctor, F. Ding, N. V. Dokholyan, Discrete molecular dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 80–92 (2011).

49. A. Emperador, T. Meyer, M. Orozco, United-atom discrete molecular dynamics of proteins using physics-based potentials. *J. Chem. Theory Comput.* **4**, 2001–2010 (2008).

50. N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, E. I. Shakhnovich, Discrete molecular dynamics studies of the folding of a protein-like model. *Fold. Des.* **3**, 577–587 (1998).

51. D. Shirvanyants, F. Ding, D. Tsao, S. Ramachandran, N. V. Dokholyan, Discrete molecular dynamics: An efficient and versatile simulation method for fine protein characterization. *J. Phys. Chem. B* **116**, 8375–8382 (2012).

52. D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).

53. G. Bussi, A. Laio, Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2**, 200–212 (2020).

54. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566 (2002).

55. J. J. G. Tesmer, C. W. Dessauer, R. K. Sunahara, L. D. Murray, R. A. Johnson, A. G. Gilman, S. R. Sprang, Molecular basis for P-site inhibition of adenylyl cyclase. *Biochemistry* **39**, 14464–14471 (2000).

56. T. C. Mou, A. Gille, S. Suryanarayana, M. Richter, R. Seifert, S. R. Sprang, Broad specificity of mammalian adenylyl cyclase for interaction with 2′,3′-substituted purine- and pyrimidine nucleotide inhibitors. *Mol. Pharmacol.* **70**, 878–886 (2006).

57. T. C. Mou, N. Masada, D. M. F. Cooper, S. R. Sprang, Structural basis for inhibition of mammalian adenylyl cyclase by calcium. *Biochemistry* **48**, 3387–3397 (2009).

58. A. Sethi, J. Eargle, A. A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6620–6625 (2009).

59. R. B. Fenwick, L. Orellana, S. Esteban-Martín, M. Orozco, X. Salvatella, Correlated motions are a fundamental property of β-sheets. *Nat. Commun.* **5**, 4070 (2014).

60. G. M. Süel, S. W. Lockless, M. A. Wall, R. Ranganathan, Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59–69 (2003).

61. J. Monod, J. Wyman, J. P. Changeux, On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).

62. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

63. S. C. Van Keulen, U. Rothlisberger, Effect of N-terminal myristoylation on the active conformation of $G\alpha_{i1}$-GTP. *Biochemistry* **56**, 271–280 (2017).

64. M. Mirdita, L. Von Den Driesch, C. Galiez, M. J. Martin, J. Soding, M. Steinegger, Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2017).

65. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).

66. T. Giorgino, PLUMED-GUI: An environment for the interactive development of molecular dynamics analysis and biasing scripts. *Comput. Phys. Commun.* **185**, 1109–1114 (2014).

67. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

68. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

69. B. Hess, P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).

70. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

71. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

72. O. Allnér, L. Nilsson, A. Villa, Magnesium ion-water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **8**, 1493–1502 (2012).

73. K. L. Meagher, L. T. Redman, H. A. Carlson, Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **24**, 1016–1025 (2003).

74. B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).

75. The PLUMED consortium, Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).

76. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).

77. A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 20603 (2008).

78. D. Branduardi, G. Bussi, M. Parrinello, Metadynamics with adaptive gaussians. *J. Chem. Theory Comput.* **8**, 2247–2254 (2012).

79. J. Eargle, Z. Luthey-Schulten, NetworkView: 3D display and analysis of protein·rNA interaction networks. *Bioinformatics* **28**, 3000–3001 (2012).

80. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

81. G. Bussi, D. Branduardi, Free-energy calculations with metadynamics: Theory and practice. *Rev. Comput. Chem.* **28**, 1–49 (2015).

82. N. Bešker, F. L. Gervasio, *Methods in Molecular Biology* (Springer, 2012), vol. 819, pp. 501–513.

83. A. Barducci, M. Bonomi, M. Parrinello, Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 826–843 (2011).

84. L. S. M. Evenseth, R. Ocello, M. Gabrielsen, M. Masetti, M. Recanatini, I. Sylte, A. Cavalli, Exploring conformational dynamics of the extracellular *Venus flytrap* domain of the GABA_B receptor: A path-metadynamics study. *J. Chem. Inf. Model.* **60**, 2294–2303 (2020).

85. F. Colizzi, M. Masetti, M. Recanatini, A. Cavalli, Atomic-level characterization of the chain-flipping mechanism in fatty-acids biosynthesis. *J. Phys. Chem. Lett.* **7**, 2899–2904 (2016).

86. M. Bernetti, M. Masetti, M. Recanatini, R. E. Amaro, A. Cavalli, An integrated Markov state model and path metadynamics approach to characterize drug binding processes. *J. Chem. Theory Comput.* **15**, 5689–5702 (2019).

Citation: F. Colizzi, M. Orozco, Probing allosteric regulations with coevolution-driven molecular simulations. *Sci. Adv.* **7**, eabj0786 (2021).