



# Thermodynamic architecture and conformational plasticity of GPCRs

Received: 18 May 2022

Accepted: 29 December 2022

Published online: 09 January 2023

 Check for updatesSathvik Anantakrishnan<sup>1</sup> & Athi N. Naganathan<sup>1</sup>  


G-protein-coupled receptors (GPCRs) are ubiquitous integral membrane proteins involved in diverse cellular signaling processes. Here, we carry out a large-scale ensemble thermodynamic study of 45 ligand-free GPCRs employing a structure-based statistical mechanical framework. We find that multiple partially structured states co-exist in the GPCR native ensemble, with the TM helices 1, 6 and 7 displaying varied folding status, and shaping the conformational landscape. Strongly coupled residues are anisotropically distributed, accounting for only 13% of the residues, illustrating that a large number of residues are inherently dynamic. Active-state GPCRs are characterized by reduced conformational heterogeneity with altered coupling patterns distributed throughout the structural scaffold. In silico alanine-scanning mutagenesis reveals that extra- and intra-cellular faces of GPCRs are coupled thermodynamically, highlighting an exquisite structural specialization and the fluid nature of the intramolecular interaction network. The ensemble-based perturbation methodology presented here lays the foundation for understanding allosteric mechanisms and the effects of disease-causing mutations in GPCRs.

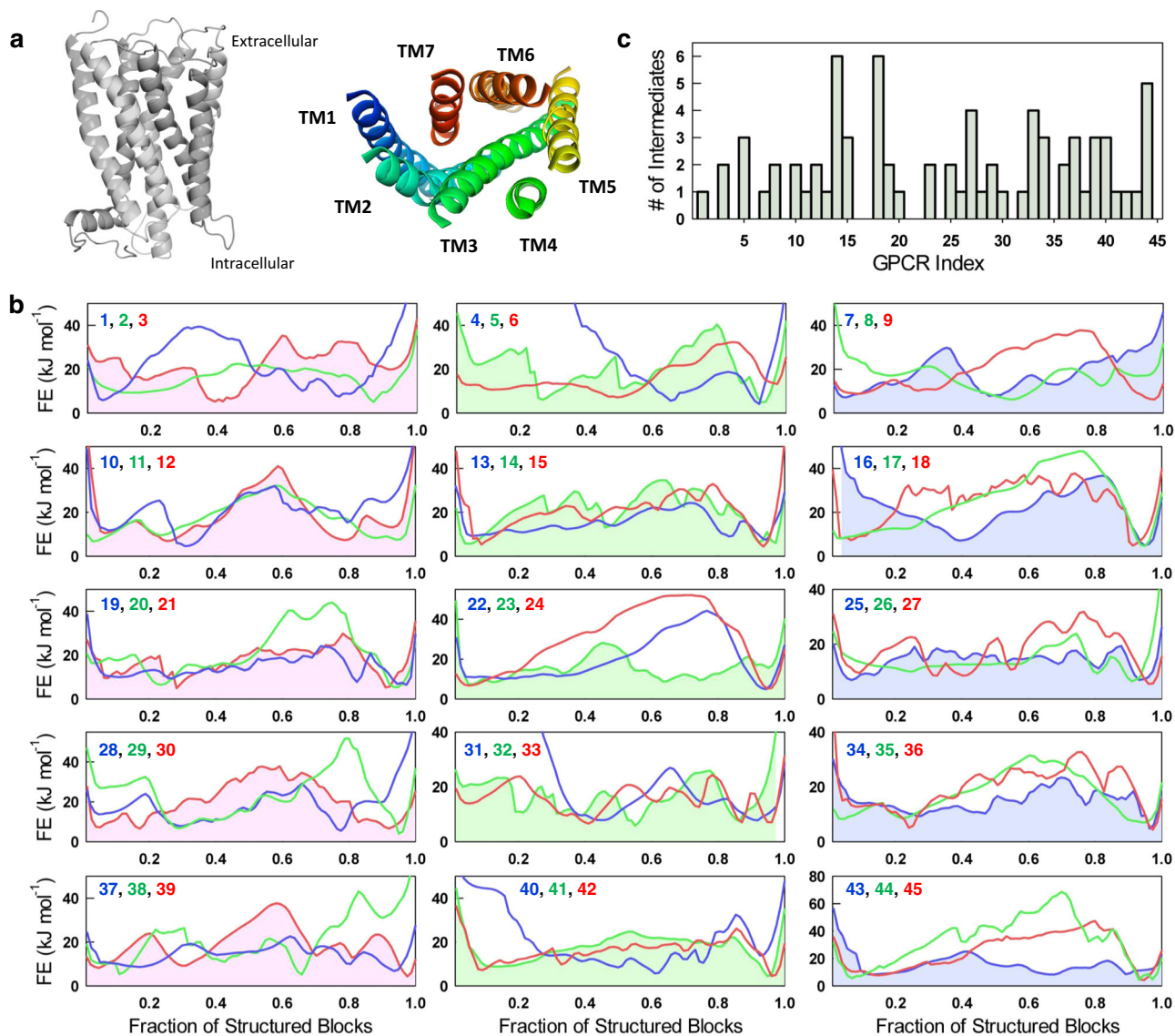
G protein-coupled receptors (GPCRs) are a large superfamily of integral membrane proteins found across the eukaryotic tree of life that are involved in numerous critical signaling processes. The human genome is known to contain over 800 different GPCRs with roles in vision, taste, smell, neurotransmission, immunoregulation, homeostasis, and growth<sup>1</sup>. Their physiological importance and the variety of processes in which they are involved are well illustrated by the fact that over 30% of clinically approved drugs target GPCRs<sup>2</sup>. Mutations in GPCRs have been implicated in a wide variety of diseases, including retinitis pigmentosa, thyroid disease, epilepsy, fertility disorders, and carcinomas<sup>3,4</sup>.

GPCRs are divided into six classes based on their functions and sequence homology, with the class A (Rhodopsin-like) receptors comprising the largest group. All GPCRs share a common transmembrane domain structure consisting of seven helices arranged in a highly conserved topology (Fig. 1a). This transmembrane domain, also called the 7TM domain, is involved in ligand binding, allosteric signal transduction, and the binding and subsequent activation of downstream

effector proteins. Ligand binding on the extracellular side of the helix bundle induces allosteric conformational changes that result in G-protein binding and activation on the intracellular side<sup>5</sup>. Signaling through GPCRs is induced by a wide variety of stimuli, including heat, mechanical stresses, small molecules, and peptides. Signals are transmitted within the cell through signaling transducers, heterotrimeric G proteins, and  $\beta$ -arrestins. GPCRs contain several highly conserved sequence motifs and structural features. The first and second extracellular loops (ECL1 and ECL2) contain conserved cysteine residues that form a disulfide bridge. On the intracellular side, many GPCRs in their inactive state contain an “ionic lock,” a network of salt bridges between residues in TM3 and TM6 (Glu6.30, Arg3.50, and Asp3.49 in  $\beta_2$ -AR using the Ballesteros–Weinstein numbering scheme)<sup>5</sup>. The helical bundle contains multiple conserved sequence motifs that act as “microswitches” which are involved in GPCR activation and stabilize the conformation of the transmembrane helices in the active state. These include the D[E]RY sequence in TM3, the NPxxY motif in TM7, and the PIF motif at the interface of TM3, TM5, and TM6<sup>6,7</sup>.

<sup>1</sup>Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India.

 e-mail: [athi@iitm.ac.in](mailto:athi@iitm.ac.in)



**Fig. 1 | GPCR free energy profiles.** **a** Representative structure of a GPCR with the seven transmembrane (TM) helices viewed side on (left) and from the extracellular side (right). **b** One-dimensional free energy (FE) profiles of the 45 GPCRs as a function of the reaction coordinate, the fraction of structured blocks, at the melting temperature. The more folded conformations populate to the right of the free energy profile (high reaction coordinate values), while the partially structured

states and unfolded conformations will be populated to the left (intermediate and low reaction coordinate values, respectively). The numbers on the top left of every panel are the GPCR indices. **c** The number of intermediates determined from the free energy profiles in panel b assuming a *IRT* threshold. Source data are provided as a Source data file.

Despite the importance of GPCRs and their ubiquitous presence in eukaryotic species, the extent of native ensemble heterogeneity in GPCRs is an open question. Their presence in the membrane makes their purification and reconstitution for biophysical experiments difficult. Further, the responses of various GPCRs to inductive stimuli take place on timescales ranging from milliseconds to hours<sup>8</sup>. Molecular dynamics (MD) simulations of the folding of these large receptors (the TM helices alone are ~300 residues in length) in their native membrane environment over long timescales are computationally challenging. Aside from all-atom MD simulations<sup>9–15</sup>, multiple biophysical techniques have been used to probe GPCR conformational dynamics, ligand binding, and GPCR–G-protein interactions at timescales of nanoseconds to seconds. These include Förster resonance energy transfer (FRET)<sup>16–20</sup>, hydrogen/deuterium exchange mass spectrometry (HDX-MS)<sup>21–25</sup>, electron paramagnetic resonance (EPR) spectroscopy<sup>26–30</sup>, and nuclear magnetic resonance (NMR) spectroscopy<sup>31–34</sup>. Alongside static structures of GPCRs in unbound, ligand-bound, and transducer-bound

states, these experiments have revealed the helix movements that occur upon GPCR activation, specific structural features that mediate these movements, and the possibility of intermediates during the transition between active and inactive conformations. Over the past decade, advancements in crystallization techniques and the design of stable GPCR fusion protein constructs have allowed for the structures of several GPCRs to be solved<sup>35</sup>. Structures of the GPCRs, however, do not provide information on ensemble features, but can serve as an excellent starting point to be used in conjunction with structure-based methods capable of investigating conformational flexibility.

In this work, we employ an Ising-like statistical mechanical model termed the Wako–Saitō–Muñoz–Eaton (WSME) model<sup>36,37</sup>, which has been quite successful in capturing the folding mechanisms and conformational landscapes of water-soluble proteins<sup>38–40</sup>, to explore the structural–thermodynamic hallmarks underlying the GPCR architecture in the ligand-free form. Despite the simplicity of the approach, we not only predict many known features of GPCRs, but also provide a

detailed view of their complex conformational landscapes, which can be used in conjunction with experiments to explore native ensemble heterogeneity, populated substates and intermediates, activation mechanisms, and allostery.

## Results

### Sequence and structural diversity in the GPCR database

Sequences corresponding to each of the 45 GPCR structures were used to generate a multiple sequence alignment (MSA) using ClustalW<sup>41</sup>. A percentage sequence similarity matrix was computed from pairwise similarities between the sequences in the MSA. Most GPCR sequences in our database exhibit low pairwise similarities, yielding a mean similarity of 9.8% ( $\sigma=2.4\%$ ) between non-identical sequences. A high sequence similarity (59.2%) is observed between the two metabotropic glutamate receptors (GPCRs 11 and 35). Similarly, several other GPCRs belonging to the same receptor subfamily display higher than average pairwise sequence similarities. These include the chemokine receptors (GPCRs 2, 4, 20, 24, 36, and 44), the  $\beta$ -adrenoceptors (GPCRs 3 and 7), the proteinase-activated receptors (GPCRs 6 and 21), the opioid receptors (GPCRs 8, 9, 10, and 19), and the melatonin receptors (GPCRs 41 and 42). The structural diversity of the GPCR dataset was also probed by computing pairwise root-mean-square deviations (RMSD) between the structures using the Dali protein structure comparison server<sup>42</sup>. Although the receptors show a high level of sequence divergence, structural similarity is found to be quite high with a mean pairwise  $C_{\alpha}$ -RMSD of 3.1 Å ( $\sigma=0.6$  Å). The only standout GPCR being the  $\beta_2$ -adrenergic receptor ( $\beta_2$ AR; GPCR7) that displays high pairwise RMSD values against all other structures used, including the  $\beta_1$ -adrenoceptor with which it shares a subfamily (RMSD  $4.8 \pm 0.6$  Å). The sequence-structure analysis effectively reveals that the dataset chosen is diverse enough to explore generic trends in the folding-conformational behaviors of GPCRs.

### Folding free-energy profiles and intermediates

The bWSME model was used to iteratively generate heat capacity curves at different values of the van der Waals (vdW) interaction energy per native contact ( $\xi$ ) while keeping every other parameter constant (Supplementary Figs. 1 and 2 and Supplementary Table 1). The magnitude of  $\xi$  that resulted in an apparent melting temperature ( $T_m$ ) of 333 K was selected (see Methods). This was done to ensure that the energy scales match the average melting temperature of mesophilic proteins, which is ~333 K. Note that the melting temperature of GPCRs is generally lower than 333 K and is expected to be different depending on the GPCR identity. The higher  $T_m$  value assumed here is to ensure that the predictions constitute the lower limit of conformational heterogeneity. The effective mean of  $\xi$  across the 45 proteins is  $-48.9 \pm 2.76$  J mol<sup>-1</sup> per native contact, indicating that none of the structures exhibit unique differences in packing that could contribute to extreme  $\xi$  values. In fact, the magnitude of  $\xi$  matches that of the 6–12 Lennard–Jones interaction potential between two carbon atoms ( $-46.1$  J mol<sup>-1</sup> at 6 Å) calculated from atomic-level force-field parameters<sup>43</sup>.

One-dimensional free-energy profiles (1D FEPs) were then generated at 333 K as a function of the fraction of structured blocks, which is a natural coordinate for the WSME model (“Methods”). The complexity of the profiles is better observed at 333 K as the favorable gradient towards the folded state at say, 298 or 310 K, obscures the features. The high sequence diversity observed in our dataset is expected to contribute to large differences among the free energy profiles and this is indeed the case. For example, some GPCRs present two-state-like free energy profiles with a large thermodynamic barrier between the folded and unfolded states (Fig. 1b). These include Free fatty acid receptor 1 (GPCR22, Supplementary Table 1) and C-C chemokine receptor type 5 (GPCR24). Others, like P2Y purinoceptor 1 (GPCR14), Orexin receptor type 1 (GPCR18), and Prostaglandin E2 receptor EP3 subtype (GPCR40), exhibit multi-state profiles containing

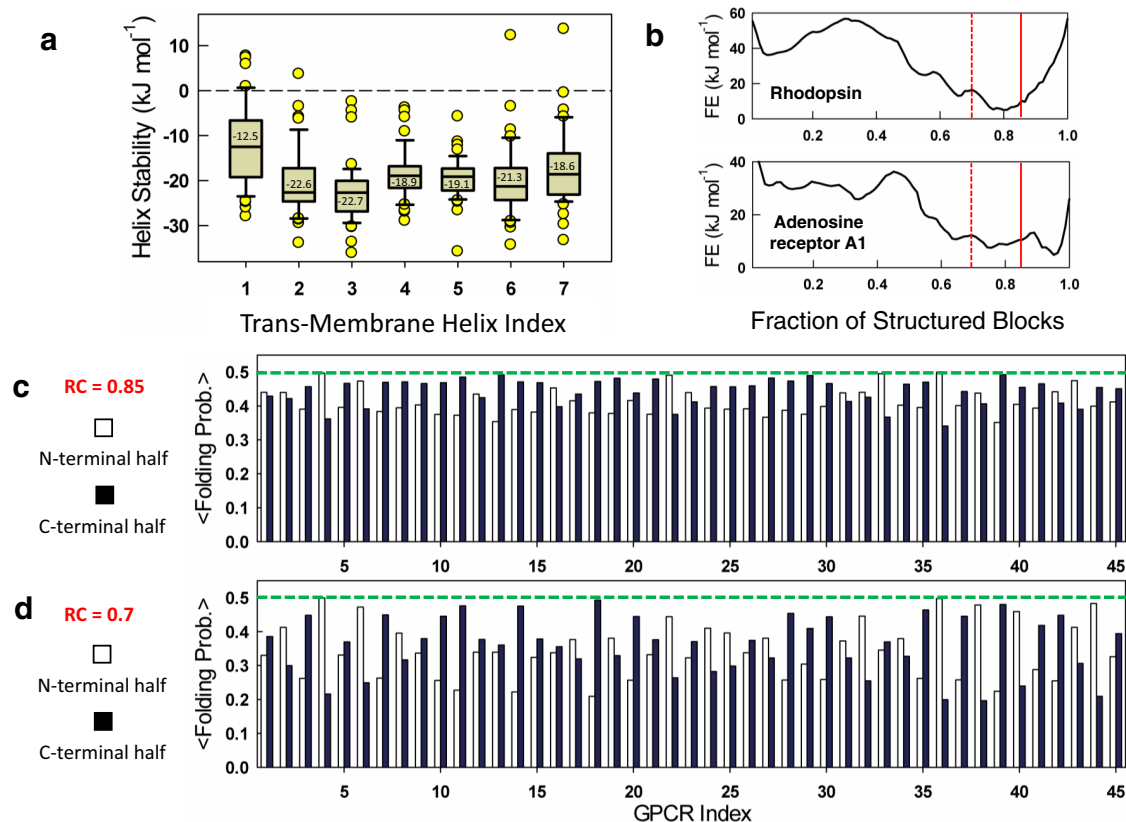
numerous intermediates. The free energy profile of Free fatty acid receptor 1 (GPCR22) features a large free energy barrier between the folded and unfolded minima and a narrow folded-state minimum. On the other hand, the free energy profile of Adenosine receptor A1 (GPCR23) features a broad folded-state minimum. Type-2 angiotensin II receptor (GPCR25), Substance-P receptor (GPCR37), and Calcitonin receptor (GPCR43) display free energy profiles that are largely flat, suggestive of a loosely coupled structural scaffold. The positions of major intermediates on the free energy profiles also differ between different GPCRs. While the free energy profile of Sphingosine 1-phosphate receptor 1 (GPCR5) features intermediates that precede the major folding barrier,  $\beta_1$  adrenergic receptor (GPCR7) populates intermediates after the major folding barrier. In many cases, the native ensemble is not defined by a single state, but by a collection of sub-states either over a barrier or as a continuum of states, and this can be seen in Rhodopsin, P2Y purinoceptor, Neurotensin receptor type 1, Cannabinoid receptor 2, and Thromboxane A2 receptor.

We estimate the number of intermediates *via* a simple heuristic: a local minimum on the 1D FEP is considered to be a partially structured intermediate if it was separated from its neighboring minima by free energy barriers of at least 1 *RT*. According to this criterion, the 1D FEPs of most GPCRs in the dataset are found to contain at least 2–3 intermediates (Fig. 1c and Supplementary Table 2). P2Y purinoceptor 1 (GPCR14) and Orexin receptor type 1 (GPCR18) populate the highest number of intermediates (6). It is important to note that the observed heterogeneity in the 1D FEP is only a lower limit. This is because the reaction coordinate, the number of structured blocks, lumps together millions of microstates to construct partial partition functions and hence folding free energy profiles.

Though the reliability of the WSME model predictions has been validated in numerous water-soluble proteins, it is not clear if they are equally applicable to membrane-associated systems. The robustness of the bWSME model energy function is showcased by studying two bacterial membrane proteins, GlpG and PagP (Supplementary Table 3), whose folding mechanisms have been re-constructed from experiments. GlpG, which consists of six transmembrane  $\alpha$ -helices, folds primarily through a mechanism that involves the folding of the entire N-terminal region of the protein before the C-terminal region folds (Supplementary Fig. 3a–d). This is in close agreement with data from single-molecule unfolding experiments and mutational analysis, which indicate that the C-terminal region is more unstable and the presence of an N-terminal biased folding nucleus<sup>44</sup>. The  $\beta$ -barrel membrane protein PagP folds via an intermediate in which parts of both the N- and C-terminal regions are structured, with a higher probability of C-terminal structure<sup>45</sup>. This observation is again in good agreement with our model, which yields a two-dimensional free energy surface with a significant local minimum in which both the N- and C-terminal regions are partially structured with a higher structural disposition towards the C-terminal strands (Supplementary Fig. 3e–h). The agreement of the model predictions with experimentally constructed folding mechanisms thus attests to the robustness of our method and the uniform dielectric constant employed for studying membrane proteins. We delve into the thermodynamic architecture of GPCRs in the sections below.

### Helix stabilities and conformational plasticity

To examine how stability determinants are distributed across the GPCR structures, the folding probability of every residue in the protein was calculated at 310 K by summing up the statistical weights of microstates with a specific residue folded and their relative contribution to the total partition function (“Methods” and Eqs. 6–8). These residue folding probabilities were then used to compute the average stability of residues within each of the seven transmembrane helices for all 45 GPCRs (Fig. 2a). Note that this calculation allows for the estimation of the helix stability in the context of the structure



**Fig. 2 | Helix stabilities and conformational plasticity at 310 K.** **a** Box plot of individual TM helix stabilities ( $n = 45$ ). Box plots include the median line (median value indicated), the box denotes the interquartile range (IQR), whiskers denote the rest of the data distribution, and outliers are denoted by points greater than  $\pm 1.5 \times \text{IQR}$ . **b** One-dimensional free energy profiles of two representative GPCRs. The vertical dashed lines signal the reaction coordinate (RC) values of 0.85

(continuous red) and 0.7 (dashed red), respectively. **c, d** Protein regions that unfold first as one moves from the right to the left on the free energy profiles are illustrated by dividing the GPCR structures into the N-terminal half (TM helices 1–3) and C-terminal half (TM helices 4–7). **c, d** plot the means folding probability of the N- and C-terminal halves of the structure at the indicated reaction coordinate values. Source data are provided as a Source data file.

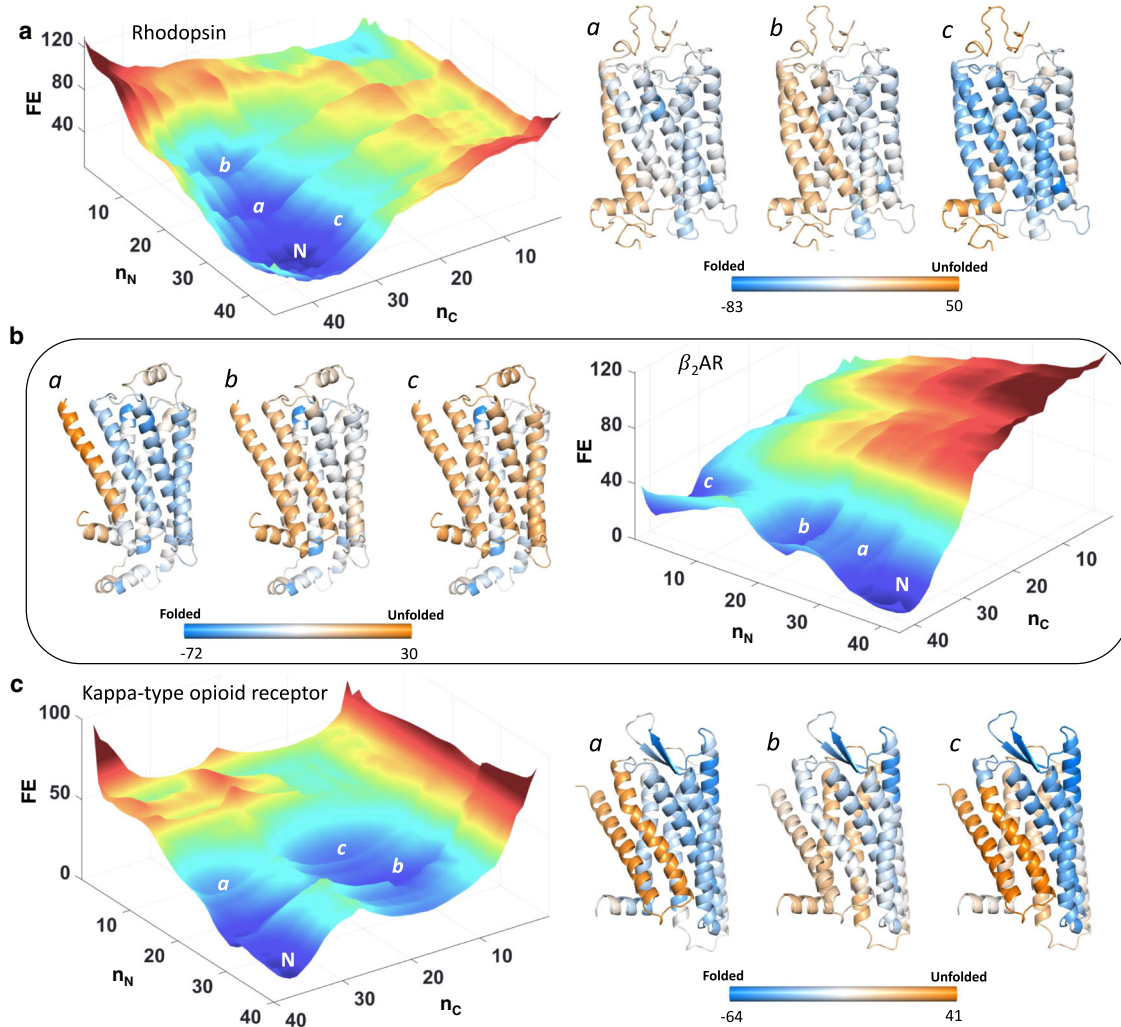
considered and not in isolation. We find that TM3 is the most stable of all the helices, with TM1 being the least stable. Thus, without explicitly considering the disulfide bridge between the first and second extracellular loops (ECL1 and ECL2), the model is still able to predict the larger stability of TM3. The stabilities of TM1, 6, and 7 vary substantially in the dataset studied, and in some GPCRs, these helices are unstructured even in the native ensemble (positive helix stability values in Fig. 2a).

As a second step, one-dimensional free energy profiles were constructed at 310 K as a function of the reaction coordinate, the number of structured blocks (Fig. 2b). Regions of the GPCR that unfold first or are partially structured in the native ensemble are identified by choosing two specific regions on the reaction coordinate (RC = 0.85 and RC = 0.7; vertical lines in Fig. 2b) and plotting the probability of structure in the N-terminal half ( $\langle p_{F,N} \rangle$ ) versus C-terminal half of the structure ( $\langle p_{F,C} \rangle$ ) (Fig. 2c, d). The former accounts for the first three TM helices while the latter accounts for the TM helices 4–7. At RC = 0.85 that corresponds to the near-fully folded native ensemble, both the N- and C-terminal halves are already partially structured in the majority of GPCRs with  $\langle p_{F,N} \rangle$  of 0.41 compared to  $\langle p_{F,C} \rangle$  of 0.44 on average (Fig. 2c). Importantly, 30 of the 45 GPCRs exhibit more unfolding in the N-terminal half compared to the C-terminal half. Minor perturbations can be mimicked by observing the stability patterns at RC = 0.7 where the protein is marginally more destabilized (Fig. 2d). Under these conditions, both the protein halves are similarly unfolded on average across all proteins ( $\langle p_{F,N} \rangle$  of 0.34, and  $\langle p_{F,C} \rangle$  of 0.36), with the distribution flipped in favor of more unstructured C-terminal halves in 27 GPCRs (Fig. 2d). Given that TM1, 6, and 7 exhibit lower stabilities

compared to other helices, it is likely that these regions sample partially structured states in the native ensemble. Topologically, this conformational behavior is expected as TM1 is the most weakly packed of all helices, interacting only with TM2 in most proteins and also with TM7 in some (Fig. 1a). On the other hand, TM7 is relatively more packed, directly interacting with all helices except TMs 4 and 5, and hence it is less likely to sample unstructured states compared to TM1.

To investigate this further, 2D free energy landscapes were generated for all GPCRs at 310 K, with the number of structured blocks at the N- or C-terminal region as coordinates. Such a 2D landscape has been particularly successful in capturing functionally relevant sub-states in multiple large water-soluble proteins<sup>40,46,47</sup>. For example, consider the free-energy landscapes of Rhodopsin (GPCR1), the  $\beta_2$ AR (GPCR3) and the Kappa-type opioid receptor (GPCR8) at 310 K (Fig. 3). The native ensembles are broad in the GPCRs considered, but with differences in the extent and nature of conformations populated. The Rhodopsin free energy surface (Fig. 3a) is indicative of a continuum of states in the native basin (states *a* and *b*), while the states *b* and *c* in the  $\beta_2$ AR (Fig. 3b) and all labeled states in the Kappa-type opioid receptor (Fig. 3c) are intermediate-like, and are populated over a marginal thermodynamic barrier. The partially structured states *c* and *a* in Rhodopsin and the  $\beta_2$ AR, respectively, will however not have a large residence time as they appear as “excited states” along the coordinate (they do not constitute a minima on the landscape). In these three GPCRs, it is TM1 that exhibits the largest degree of unfolding.

In Rhodopsin state *a* (Fig. 3a), one would expect the unfolding of TM1 to not affect the adjacent helices, but it is clear that the free energy of folding (Eq. 7) of almost all the helices are perturbed—they should



**Fig. 3 | GPCR conformational landscapes and native ensemble heterogeneity.** Free energy landscapes of Rhodopsin (a),  $\beta_2$ AR (b), and Kappa-type opioid receptor (c). The free energy values and the color bars are in units of  $\text{kJ mol}^{-1}$ . The structures are color coded according to the color bars with light blue and dark orange representing fully folded and fully unfolded conformational status, respectively.

Intermediate free energy values colored in white represent regions of intermediate stability. The major conformational states *a*, *b*, and *c* are shown adjacent to the free energy landscapes with *N* representing the folded native ensemble. Source data are provided as a Source data file.

be in the dark blue color range (more folded) but instead fall in the region between cyan and white (partially unfolded). This is a consequence of the fact that a loss of interactions between TM1 and TMs 2 and 7 in turn destabilizes the TMs adjacent to them but to a lesser extent, similar to the effect of mutations on protein structure<sup>48</sup>. State *c* in Rhodopsin is characterized by fully folded TMs 1–4 while TMs 5–7 are partially structured. In the  $\beta_2$ AR, unstructured TMs 1–2 are the predominant substates (states *a* and *b* in Fig. 3b), similar to the state *a* in the kappa-type opioid receptor (Fig. 3c). Additionally, the substate *a* in  $\beta_2$ adrenergic receptor exhibits partial structure in TMs 6 and 7 (white in Fig. 3b), which mirrors experimental observation of substates involving significant mobility in the same set of helices<sup>49</sup>. Furthermore, partial structure in TMs 1, 2, 6, and 7 of the kappa-type opioid receptor promote the population of an intermediate *c* that has only TMs 3, 4, and 5 folded (Fig. 3c). To summarize, it appears that while partial unfolding of TM1 is a dominant substate in GPCRs, there can be substantial variation in the nature of the states populated and their relative populations.

#### Anisotropic distribution of coupling free energy magnitudes

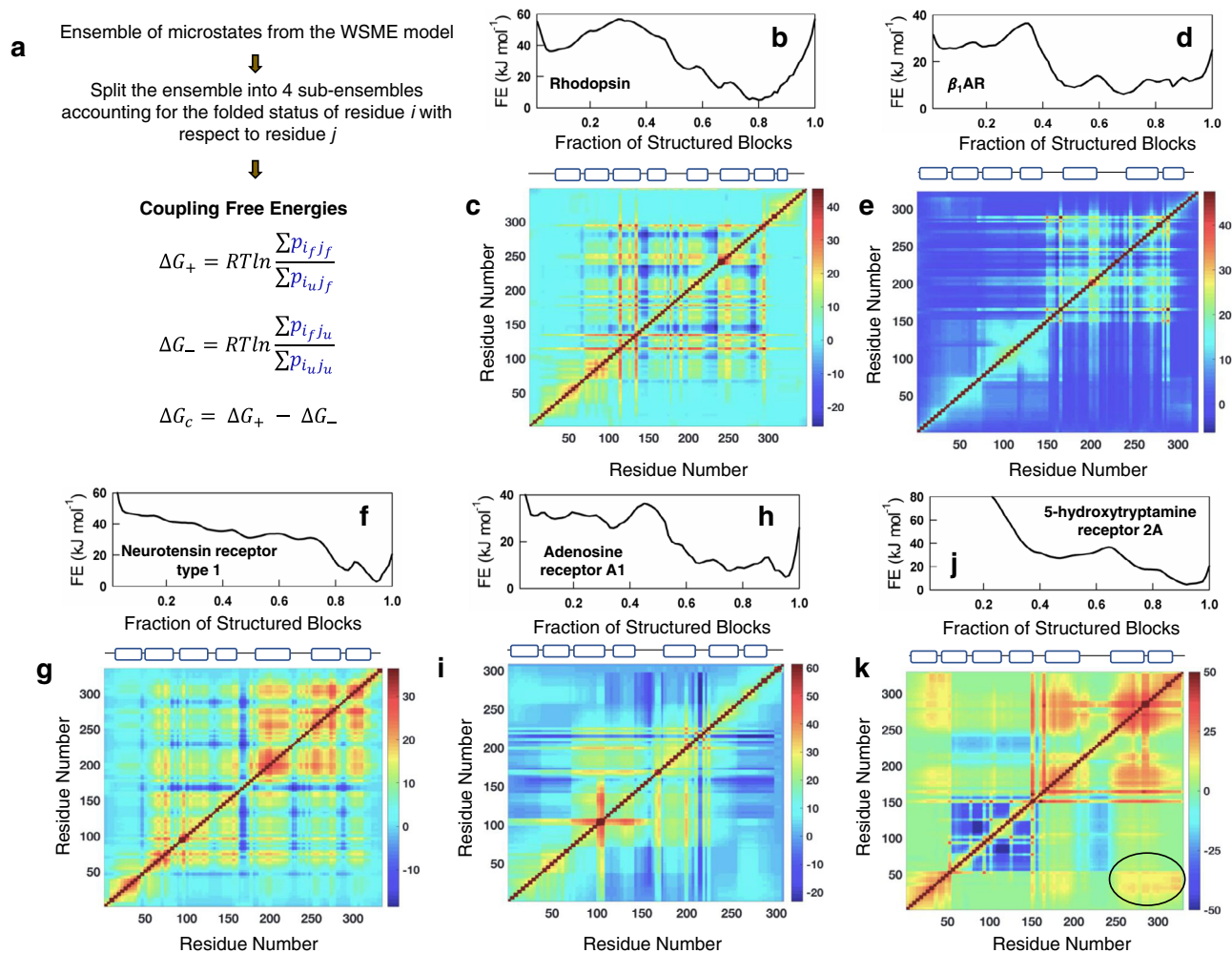
GPCR activation mechanisms are dependent not just on the thermodynamic stabilities of individual helices (in the structural context) but

also the extent to which these stabilities are modulated via altered structural patterns and contacts *between* helices on ligand binding. A precise understanding of this could be gleaned by computing the extent to which the different regions of the protein are thermodynamically coupled to each other<sup>50,51</sup>. We calculate coupling free energies between residues from the bWSME model by grouping the ensemble of microstates into four different sub-ensembles for every residue *i* (Fig. 4a):  $\sum p_{ijf}$  sums over the probabilities of all states in which both residues *i* and *j* are folded,  $\sum p_{iju}$  sums over probabilities of states in which residue *i* is folded and *j* is unfolded, and similarly for  $\sum p_{ijf}$  and  $\sum p_{iju}$ <sup>39</sup>. From these groupings, one could calculate positive ( $\Delta G_+$ ), negative ( $\Delta G_-$ ) and effective ( $\Delta G_c$ ) coupling free energies<sup>39,52</sup> between different residues using:

$$\Delta G_+ = RT \ln \left( \frac{\sum p_{ijf}}{\sum p_{ijf}} \right) \text{ and } \Delta G_- = RT \ln \left( \frac{\sum p_{iju}}{\sum p_{iju}} \right) \quad (1)$$

$$\Delta G_c = \Delta G_+ - \Delta G_-$$

Positive coupling free energies quantify the extent to which residues *i* and *j* are coupled via direct interactions or through long-range



**Fig. 4 | Thermodynamic architecture of GPCRs.** **a** A flow chart describing the methodology employed to calculate coupling free energies from the microstate populations. **b, d, f, h, j** One-dimensional free energy profiles at 310 K. **c, e, g, i, k** The corresponding effective coupling free energy ( $\Delta G_c$ ) matrices. The color bars are in the spectral scale and in units of  $\text{kJ mol}^{-1}$ , with dark red and dark

blue representing strong and weakly coupled residues, respectively. The cartoons on top indicate the position of the secondary structure elements along the sequence. The oval in **k** highlights that the TM helices 1 and 7 are strongly coupled in 5-hydroxytryptamine receptor 2A. Source data are provided as a Source data file.

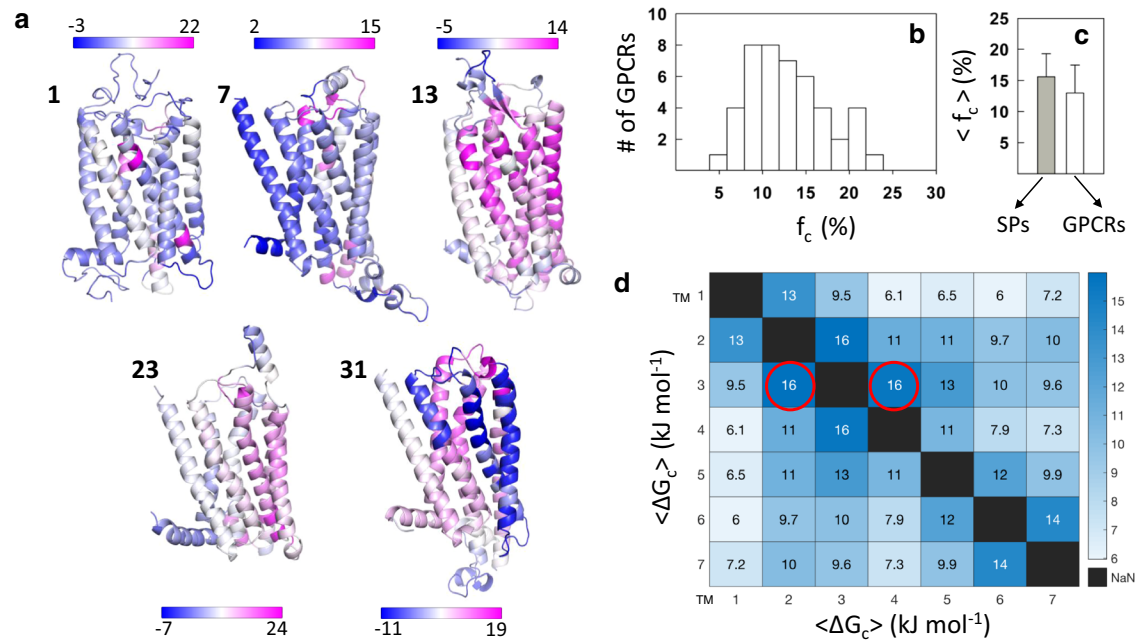
interactions in the native ensemble while the negative coupling free energies quantify the extent to which lack of spatial proximity, unfavorable interactions or large conformational entropy decouples specific structural regions from others. The balance between the two terms results in effective coupling free energies—residues that present lower effective coupling free energies are typically located in functional or dynamic regions of the structure as shown for multiple proteins in a recent work<sup>39</sup>. Importantly, coupling free energies can be calculated for every residue with respect to every other residue (and hence a square matrix can be constructed), revealing insights into the distribution of stabilization free energies in the structure.

Given the range of GPCR free energy profiles and individual helix stabilities, the coupling maps are expectedly not uniform across the GPCR dataset. For instance, in Rhodopsin (GPCR1), the broad native well in Fig. 4b is a manifestation of minimal coupling between N- and C-terminal regions (note the cyan scale in Fig. 4c). In the  $\beta_1$ AR (GPCR7), the native ensemble is not composed a single state but a continuum of conformations (Fig. 4d) which results from weak inter-residue coupling between the majority of residues in the protein (sea of blue in Fig. 4e). More complex patterns are also evident for Neurotensin receptor type 1 (Fig. 4f, g) and Adenosine receptor A1 (Fig. 4h, i), with distinct coupling free energy patterns. 5-hydroxytryptamine receptor 2A (GPCR31), meanwhile, exhibits strong coupling between residues in

its C-terminal region with particularly strong inter-helical coupling between TM1 and TM7 not seen in the other members discussed here (circled regions in Fig. 4k).

The diverse patterns in Fig. 4 are better observed by mapping the residue-averaged coupling free energies ( $\langle \Delta G_c \rangle$ , i.e., averaging along the dimensions of the symmetric matrices in Fig. 4) onto the three-dimensional structure. It can be seen that the coupled residues (different shades of magenta) are not uniformly distributed throughout the structure but are localized to specific regions in the protein (Fig. 5a). The  $\langle \Delta G_c \rangle$  values were Z-scored to account for intrinsic differences in the range of coupling free energies, and residues that exhibit a Z-score greater than one were labeled as strongly coupled. The fraction of strongly coupled residues ( $f_c$ ) thus calculated vary for the GPCRs, but are constrained within 25% (Fig. 5b). In fact, an analysis of 25 water-soluble proteins (SPs) found that they exhibit a mean  $f_c$  of  $-16\%$ <sup>39</sup>, while the GPCR dataset exhibits a marginally lower value of  $-13\%$  (Fig. 5c). Thus, despite the membrane bound nature of GPCRs, their thermodynamic architectures do not deviate significantly from those of water-soluble proteins. The anisotropic distribution patterns of strongly coupled residues are also consistent with the observations in SPs<sup>39</sup>.

To derive more generalized inferences on which regions of the proteins are more coupled than others, the average pairwise coupling



**Fig. 5 | Anisotropic distribution of coupling free energies.** **a** The mean effective coupling free energy from the matrices in Fig. 4 are mapped on the structure of GPCRs. The GPCR indices are provided for reference. Color bars are in units of  $\text{kJ mol}^{-1}$ . Regions of the protein colored in dark magenta represent strongly coupled residues while those in dark blue are not coupled. **b** Distribution of the fraction of strongly coupled residues ( $f_c$ ) in GPCRs ( $n = 45$ ). **c** Comparison of the mean fraction of strongly coupled residues in GPCRs ( $n = 45$ ;  $13.0 \pm 4.5$ ) and soluble proteins (SPs,

$n = 25$ ;  $15.7 \pm 3.7$ ). Data are presented as mean values  $\pm$  one standard deviation. **d** Effective coupling free energies between transmembrane (TM) helices averaged over the 45 GPCRs. The mean standard deviations in  $\langle \Delta G_c \rangle$  are of the order of  $4\text{--}6 \text{ kJ mol}^{-1}$ , indicating significant variability. Despite this variability, TM3 turns out to be strongly coupled to every other TM helix. Source data are provided as a Source data file.

between helices across all 45 inactive GPCR structures is calculated to construct the inter-TM coupling matrix (Fig. 5d). Transmembrane helices located adjacent to one another are strongly coupled due to the nearest neighbor effects. TM1 and TM7, on the other hand, are weakly coupled to the rest of the structure (except to TM2 and TM5, respectively), as they constitute the termini of the protein. In agreement with Fig. 2a, TM3 is the most strongly coupled of the helices, consistent with TM3's theorized role as a structural hub that maintains the GPCR scaffold<sup>53</sup>. Furthermore, in the topological arrangement of helices, TM3 interacts with all the other helices except TM1, making this region in particular more stable and crucial to the stability and functioning of GPCRs. TM4 is strongly coupled to TM helices 2, 3, and 5, while being marginally coupled to the other helices. One standout message from the inter-TM coupling matrix is the fact that every TM is either weakly or strongly coupled to one another thermodynamically. Any modulation of the coupling free energies between a pair of helices, say by ligand binding, will necessarily affect the coupling free energies throughout the structure (vide infra).

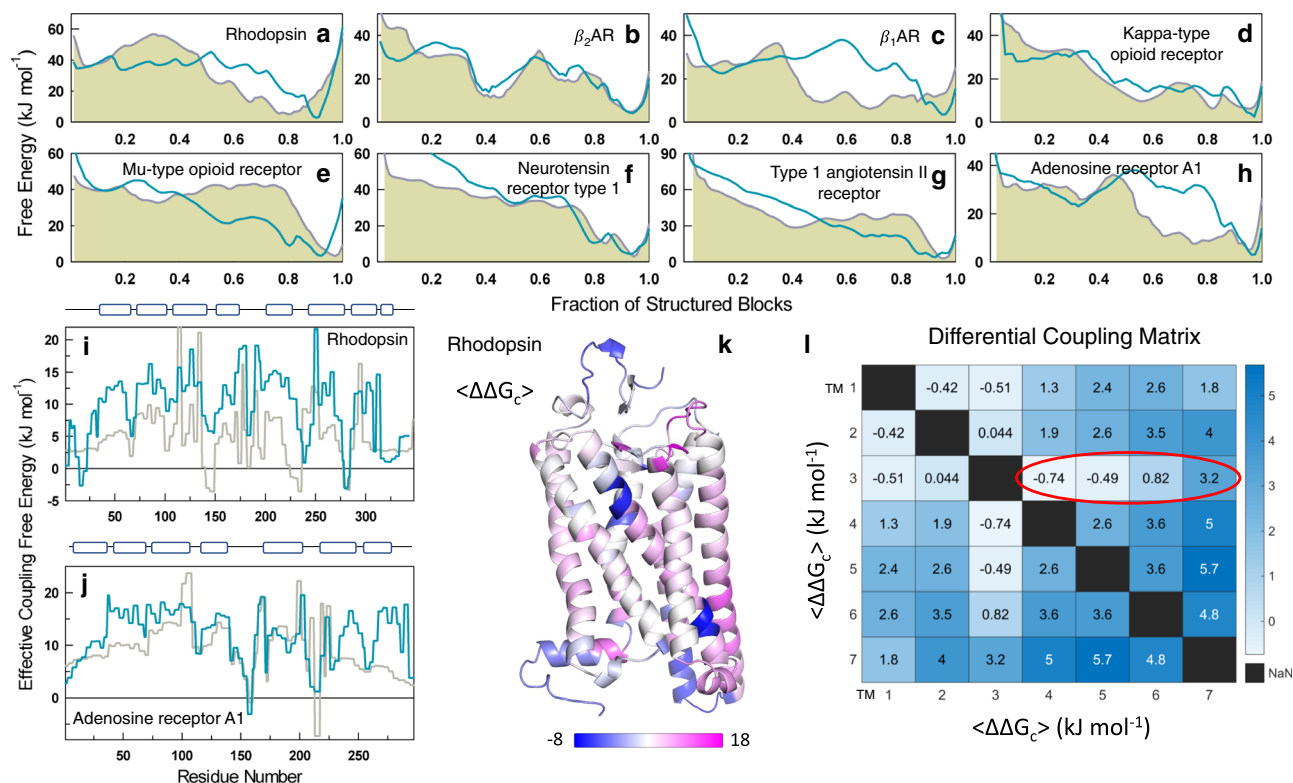
### Are active state structures more strongly coupled?

GPCR activation is characterized by large-scale movements of transmembrane helices<sup>5,14,15</sup>. In particular, activation causes TM6 to swing outward while TM5 and TM7 move in towards the helical bundle. This should affect the coupling free energies and hence the free energy profiles depending on the extent to which interactions are formed or broken between residues in TM3, TM5, TM6, and TM7. To understand this quantitatively, free energy profiles were generated for 8 GPCRs whose structures are available in both active and inactive states (Fig. 6a–h and Supplementary Table 4). The number of accessible states in the active state native ensemble are minimized in Rhodopsin (GPCR1),  $\beta_1$ AR (GPCR7), Kappa-type opioid receptor (GPCR8) and Adenosine receptor A1 (GPCR23), i.e. the native ensemble is sharper with a narrow minimum, while no significant modulations are

observed in  $\beta_2$ AR (GPCR3) and Neurotensin receptor type 1 (GPCR13). These results are consistent with the idea that inactive GPCRs are capable of sampling a variety of conformations and that agonist binding stabilizes the active-like conformation<sup>5</sup>. Particularly, the finding that  $\beta_2$ AR samples similar set of conformations in the inactive and active states is in agreement with detailed NMR experiments<sup>54</sup>. On the other hand, Mu-type opioid receptor (GPCR9) and Type 1 angiotensin II receptor (GPCR16) display a broader native ensemble in their active state, indicating that the connection between the active state and a narrower ensemble is not generalizable, at least from the perspective of 1D free energy profiles.

We further computed the effective coupling free energy matrices for the active and inactive structures and averaged them along one dimension to plot them as a function of sequence index. Residues tend to be more strongly coupled to the rest of the structure in active-state structures compared to inactive-state structures on average (Fig. 6i, j and Supplementary Fig. 4). The stronger coupling in the active form is more evident in the case of Rhodopsin (Fig. 6i), Adenosine receptor A1,  $\beta_1$ AR, and Mu-type opioid receptor (Supplementary Fig. 4) wherein nearly all helices are stabilized.  $\beta_2$ AR, on the other hand, displays little change in the degree of coupling though differences can be observed between and including TMs 5 and 6. The inactive structure is, however, more coupled in the Neurotensin receptor type 1 and Type 1 Angiotensin II receptor. The differential wiring of the interaction network in each of the protein potentially contributes to the differences we observe from the perspective of the coupling free energies, highlighting the intrinsic malleability of the contact-network in GPCRs.

Finally, the difference between the effective pairwise coupling between transmembrane helices in active and inactive structures was computed for the 8 GPCRs to generate the differential coupling matrix ( $\Delta G_{c,\text{active}} - \Delta G_{c,\text{inactive}}$ ; Fig. 6j). The mean pairwise thermodynamic coupling between most helices increases upon GPCR activation, despite the averaging across 8 GPCRs. The standard deviations are



**Fig. 6 | Active versus inactive states.** **a–h** Free energy profiles derived from the inactive (gray filled) and active (dark cyan) GPCR structures at 310 K. **i, j** Effective free energies as a function of residue index for two representative GPCRs. Gray and dark cyan represent the effective coupling free energies calculated from inactive and active states, respectively. The cartoons on top indicate the position of the secondary structure elements along the sequence. **k** Differences in effective coupling free energy (active minus inactive) from **i** is mapped on to the structure of

Rhodopsin. **l** The differential coupling matrix (active minus inactive) from the perspective of transmembrane helices. The dark red oval represents a gradation in the coupling of TM3 with the other helices on activation; this highlights the motions of other helices around TM3. The mean standard deviations in  $\langle \Delta \Delta G_c \rangle$  are of the order of 4.5–6.5  $\text{kJ mol}^{-1}$ , suggestive of large variations across the 8 pairs of active-inactive states. Source data are provided as a Source data file.

quite large, however, indicating that no two activated structures contribute to similar changes in the coupling free energies. Despite this, TM3, the most stable and most strongly coupled of the helices in the inactive state, exhibits a gradation in the coupling differences between active and inactive state structures. While it is more strongly coupled to distant helices (TM6 and TM7) in the active state, coupling with TM1, TM4, and TM5 decreases upon GPCR activation.

### Alanine-scanning reveals long-distance thermodynamic connectivity in Rhodopsin

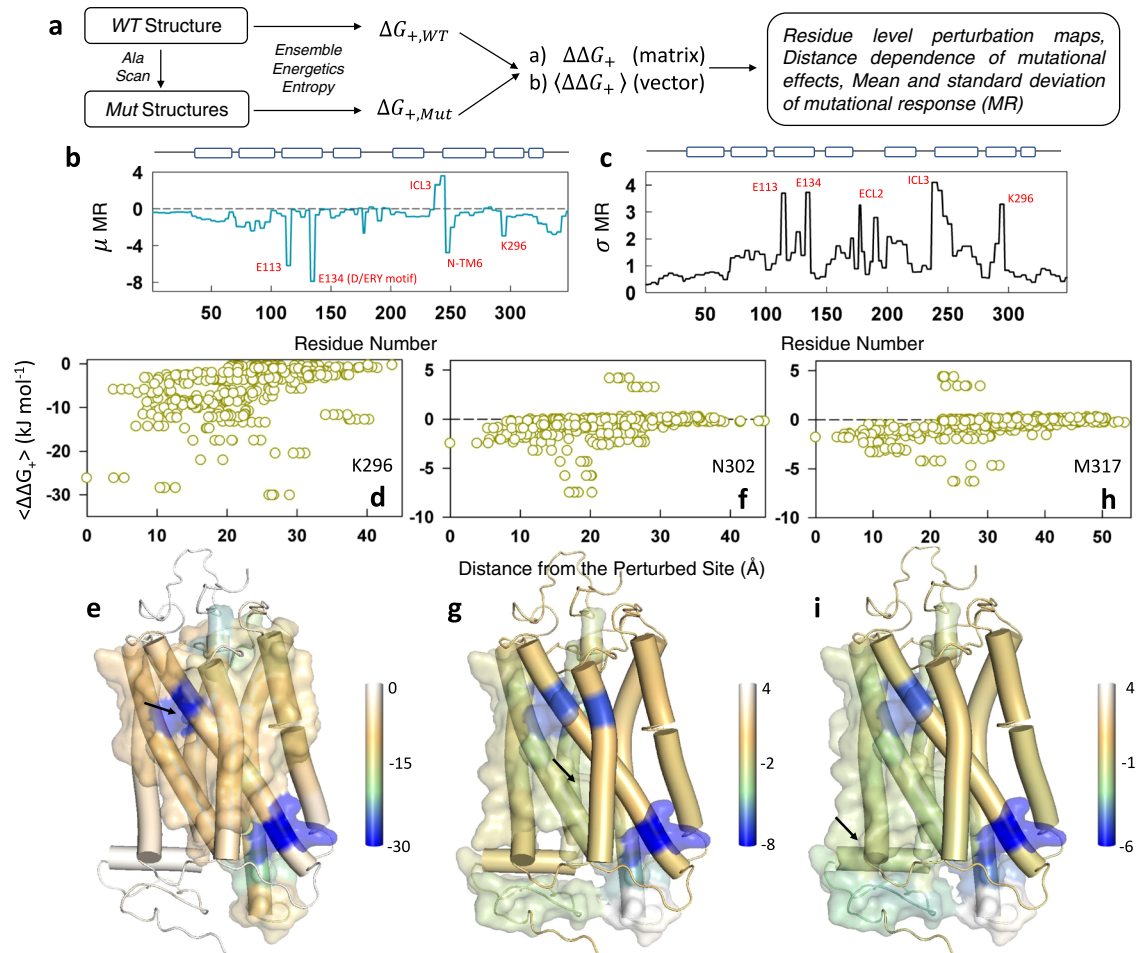
The comparison of active-inactive structures shows that the perturbations induced by activation can be pervasive and modulate long-range structural features. The extent to which the binding of a ligand influences a distant site could be potentially studied for every GPCR in the presence of agonists and antagonists. However, the WSME model does not include the atomic detail necessary for detailed modeling of subtle effects at the level of chemical interactions between the protein and the ligand, a feature that likely determines the differential effect of ligands on GPCR structures. Moreover, the model cannot be employed to reproduce the experimental protein–ligand dissociation constant (as of now), which is a necessary first step towards understanding ligand binding effects.

An alternative is to explore the extent to which every residue is coupled to every other residue in the folded conformational ensemble by performing alanine-scanning mutagenesis. Alanine substitutions at different positions will affect the interaction network to different levels depending on the immediate environment of the mutated residue, and in comparison with the WT connectivity matrix, one can extract the extent of coupling to a distant site. We consider only the positive

coupling free energy ( $\Delta G_+$ ) for this calculation, as it carries information on the states that harbor coupled residues in the native ensemble, while not considering the decoupled residues or microstates in the unfolded ensemble. For every mutant, a  $\Delta G_{+, \text{Mut}}$  matrix ( $N \times N$  matrix where  $N$  is the number of residues in the protein) is generated and referenced to the WT matrix ( $\Delta G_{+, \text{WT}}$ ) to arrive at the differential positive coupling matrix ( $\Delta \Delta G_+$ ) (Supplementary Fig. 5), which is averaged across all pairwise sites to generate the vector ( $\Delta \Delta G_+$ ) (dimension  $N \times 1$ ) (Fig. 7a). The latter carries information on the extent to which every residue is perturbed including the direction of perturbation—positive and negative change are indicative of stronger and weaker coupling, respectively—for a given alanine mutation. If the alanine mutation is performed across  $m$  sites on the protein, the resulting  $\langle \Delta \Delta G_+ \rangle$  matrix (dimension  $N \times m$ ) is employed to generate the mean  $\mu$  and standard deviation  $\sigma$  of the mutational response (MR, dimension  $N \times 1$ ).

We carried out a large-scale in silico alanine scanning mutagenesis of Rhodopsin involving 276 sites on the protein excluding the positions containing alanine, glycine and proline. The process involved introduction of mutations using PyMol<sup>55</sup>, construction of ensembles with parameters identical to that of the WT, and, following this, generation of  $\langle \Delta \Delta G_+ \rangle$  matrices (dimension  $348 \times 276$ ; Supplementary Fig. 6). The resulting mean mutational response highlights specific protein regions whose coupling magnitudes change the most on perturbations (Fig. 7b). First, the retinal-binding site residues E113 (which is a part of EGFF sequence block) and K296 (part of the FFAK block), both of which line the orthosteric site, stand out as positions that exhibit large changes upon mutational perturbations. Second, the G-protein binding pocket involving the D(E)RY motif (E113) and the





**Fig. 7 | Alanine-scanning reveals long-range communication in Rhodopsin.**

**a** Protocol employed for *in silico* alanine-scanning mutagenesis studies from the perspective of the WSME model. **b, c** Mean (**b**) and standard deviation (**c**) of the mutational response (MR in  $\text{kJ mol}^{-1}$ ). Specific residues and structural regions are highlighted for discussion in the main text. E113 and K296 are in the retinal binding pocket while E134 forms a part of the conserved D(E)RY motif that interacts with the G-protein. N-TM6 represents the N-terminal region of TM6. **d, e** Perturbation

response of K296 quantified by  $\langle\Delta\Delta G_+\rangle$  as a function of  $C_\alpha$ -distance from the mutated site (**d**).  $\langle\Delta\Delta G_+\rangle$  values are mapped on to the structure in **e** following the color code shown to the right. Residues that exhibit a  $\langle\Delta\Delta G_+\rangle$  value stronger than the average are represented in the surface mode. Arrow points to the location of the residue that is perturbed. **f-i** Same as **d, e** for N302 (**f, g**) and M317 (**h, i**), respectively. Source data are provided as a Source data file.

intracellular loop 3 (ICL3)/N-terminal region of TM6 exhibits high sensitivity to mutations across the structure. The same regions additionally exhibit larger standard deviations in the mutational response, indicating that the Rhodopsin structure exhibits intrinsic differences in dynamics (and hence thermodynamic coupling) depending on the location of the perturbation (Fig. 7c).

The residue-level  $\langle\Delta\Delta G_+\rangle$  vector that is generated for every mutation carries information on the degree to which different residues are thermodynamically coupled and hence the extent (quantified in terms of distance) to which such perturbation effects are felt. As representative examples, we discuss three different residues—K296, N302, and M317—that play critical roles in the functioning of Rhodopsin, and GPCRs in general<sup>56</sup>. Perturbation of K296 in the retinal binding pocket (i.e., a K296A mutation) induces strong destabilization across the structure when compared to the WT. This can be observed as negative  $\langle\Delta\Delta G_+\rangle$  and that modulates the folding status of residues located as far as 35–40 Å from the mutated site with the major effect within 25 Å (Fig. 7d and Supplementary Fig. 5). Mapping these magnitudes on to the structure (Fig. 7e) it is clear that any perturbation of K296 or residues around it will naturally modulate the extent of coupling at the G-protein binding site located at the intracellular side of Rhodopsin (with residues spanning TM helices 3, 5, 6, and 7). This can be seen from the surface representation for

strongly coupled residues in Fig. 7e that spans the entire length of Rhodopsin. Thus, it appears that the ligand binding pocket is strongly and thermodynamically connected to the G-protein binding pocket. Perturbation of N302 in TM7 (from the NPxxY motif) reveals that this residue is thermodynamically coupled to a majority of residues at the intracellular side (Fig. 7f, g). Remarkably, this connection, as represented by the surface map in Fig. 7g, extends all the way to the ligand-binding pocket though the magnitude of this coupling is ~4 times lower compared to K296. The residues in helix 8 have been proposed to also interact with G-proteins to enable the formation of functional complexes<sup>57</sup>. While perturbation of K296 reveals little effect on helix 8, we find that perturbation of M317 (located in helix 8) is sensed at the ligand binding site including K296 and the surrounding residues (Fig. 7h, i). The lack of reciprocity (K296 versus M317, for example) reveals that conformational modulations can be fine-tuned to accommodate a ligand with different distantly located protein regions providing their feedback to the ligand binding site(s), potentially determining their unbinding rates.

## Discussion

The diverse sequence features of the GPCR family are implicitly accounted for by the WSME model's energy-entropy function—sequence-structure-dependent conformational entropy, charge-charge

interactions and packing interactions—and the observed native ensemble heterogeneity is an emergent property of these small sequence-dependent features. One of the consistent observations is the presence of kinetic traps or intermediates in the free energy profiles; these intermediates are likely a manifestation of functional requirements as shown recently for several large water-soluble proteins<sup>40</sup>. True to this expectation, the functionally important TM helices (TMs 1, 6, and 7) are typically only marginally coupled to the rest of the structure, exhibit low intrinsic stability in the inactive state and are partially unstructured in the native ensemble. This conformational pre-equilibrium, defined as the co-existence of both fully folded and partially structured substates with varying probabilities on the conformational landscape, either over a broad native ensemble or as a series of intermediates and/or excited states, is potentially one of the reasons for the difficulty in crystallizing GPCRs. However, the precise extent of pre-equilibrium is not universally conserved—no two GPCR conformational landscapes are similar—and is likely evolutionarily selected based on the identity of the ligand and the required magnitude of functional readout. These aspects need to be studied on a case-by-case basis with appropriate experimental calibration of the model.

The magnitude of coupling free energies, which are second-order measures (unlike residue folding probabilities which are first-order measures), provides insights into the structural and thermodynamic architecture of proteins. The degree of coupling of different structural regions in GPCRs can be dramatically different despite their high structural similarity, showcasing the exquisite structural evolution driven by functional requirements. Specifically, such diversity is a consequence of the anisotropic distribution of stability patterns across the structure, with the distribution of strongly coupled residues (most strongly coupled residues are in TMs 2 and 3) and the fraction of strongly coupled residues (<30%) mirroring observations in soluble proteins. The central role of TM3 as a structural hub emerges naturally from the structural-ensemble-based calculation, in addition to the precise magnitude of coupling between different TM helices in the inactive state. Structural analysis of GPCRs based on consensus contacts has revealed extensive insights into GPCR structures, likely activation mechanisms and regions of proteins that are involved in activation<sup>5,53</sup>. We reformulate these implicitly into the WSME model and find that many active structures sample a significantly constrained conformational space. This can be explicitly observed both in the free-energy profiles and in the resulting coupling free-energy magnitudes. While it is not possible to generalize these observations to all GPCRs given the limited dataset, the mean changes in coupling free energies follow a specific pattern wherein all TM helices are more strongly coupled to the rest of the structure in the active state compared to the inactive conformation. Though TM3 remains rigid during these conformational motions, the effective coupling with adjacent helices is modulated from negative (weaker interactions) to positive (stronger interactions) in going from TM1 to TM7.

Given the large conformational flexibility in the ligand-free GPCRs, it is tempting to speculate that the observed helix mobility and partial unfolding are required for effective binding to various agonists and antagonists and for precise control of functional outcomes. In fact, simulations involving alprenolol binding to  $\beta_2AR$  point to two primary pathways involving either the channel between ECL2 and TM4/6/7 helices or between ECL2 and TM2/7<sup>58</sup>, which was also subsequently observed long time-scale MD simulations<sup>59</sup>. There is an overall consistency between simulations and the WSME model predictions. Specifically, partial structure in TMs 1/2 (state *b* in Fig. 3b) and TMs 1/2/3/4 (state *c*) can open up potential crevices and channels for ligands to bind. In Rhodopsin, retinal unbinding simulations point to unbinding from a cleft between TM 4/5 or TM5/6<sup>60</sup>. If one were to expect unbinding to be the reverse of binding from the principle of microscopic reversibility, then this requires structural flexibility and partial unfolding of the TM helices 4/5/6 and this is observable in states

*a* and *b* in Fig. 3a. The large entropic stability of inactive GPCR conformations (because of their inherently flexible nature) is therefore likely compensated by enthalpic effects through the binding of ligands in the open crevices, cavities between TM helices or in the extracellular side (agonists, antagonists, or drugs), and via G-proteins on the intracellular side.

The mechanistic basis for diseases associated with GPCR dysfunction include inactive or constitutively active receptors, under-expressed receptors, and misfolded receptors, all of which arise due to mutations distributed across the structure. It is conceivable that mutations modulate the number and nature of intermediates or many of the minor excited states, thus influencing foldability and half-life. Such modulations could appear as differential coupling patterns in the communication network and hence manifest as allosteric effects, subtly determining the binding of various agonists, antagonists and partial agonists. This communication network is extracted by performing large-scale alanine-scanning mutagenesis on Rhodopsin as a representative example. The ligand-binding pocket, closer to the extracellular face, is found to be strongly connected to the G-protein binding site at the intracellular side, as evidenced by large differences in positive coupling free energies when a mutation is introduced at the ligand binding site (Fig. 7), similar to the results of sequence-based statistical coupling analysis<sup>61</sup>. Given the large-scale connectivity map, it appears that binding of ligands is precisely coordinated by not just the binding site residues and “microswitches”, but also the folded status of many residues far from the binding pocket, a feature that likely determines the differences in affinity to agonists, antagonists and partial agonists. We would like to note that the perturbation method does not reveal the different “communication routes” nor the fluxes through them, but reports on the extent to which a distant site is perturbed and the magnitude of perturbation. Importantly, the resulting changes in positive coupling free energies are a manifestation of the differences in the underlying distribution of states (as exemplified by Eq. 1) which when mapped on to a single structure reveal distance-dependent effects. The nature and the strength of ligand binding (which is effectively a perturbation to the binding site) could therefore determine the functional output by restricting the accessible conformational space as seen in the energy landscapes of activated receptors. A similar conformational feature has been recently demonstrated in the diverse class of large nuclear receptor ligand-binding domains<sup>47</sup>, indicating that “conformational selection” and subsequent enthalpic compensation (via drug–protein contacts or GPCR–G-protein interactions) of entropic stability could be generic features underlying the energy landscapes of proteins and, hence, function.

Membranes, which are implicitly treated in the current approach as a low dielectric continuum, and their specific composition are known to affect the conformational features of GPCRs<sup>7,12</sup>. Would lipids enhance or reduce the observed heterogeneity? Since the interactions will membrane components and cholesterol will stabilize specific parts of the structure, they will likely contribute to the population of additional states with lifetimes dependent on the strength of interaction. Naturally, this would enhance the ruggedness of the conformational landscape. This expectation has already been borne out in simulations involving water-soluble proteins with non-native interactions<sup>62,63</sup>. Thus, the GPCR conformational complexity presented here is likely a lower estimate, with lipids, cholesterol, ions and pH modulations further tuning the equilibrium of states.

Finally, it is important to state that the ability of the phenomenological bWSME model to quantitatively characterize the conformational landscapes of GPCRs depends on the quality of the input structure and the kind of experimental data available for calibration. We provide a critique of our approach below to highlight potential limitations and advantages. First, the model is conventionally calibrated against heat capacity profiles for soluble globular proteins; this specific data carries information on not just the melting temperature,

but also heat capacity change, and the cooperativity of the transition which is related to the overall partition function. However, such DSC profiles or even temperature-dependent unfolding curves are challenging to measure for membrane proteins, as the lipids will themselves undergo phase transition, confounding the interpretation. The availability of such data or unfolding curves could pave the way for a sound calibration of model parameters for wild-type and mutant variants, and thus enabling quantitative predictions. Second, we consider a single low dielectric constant (value of 4) to simulate membrane proteins, as it is the simplest possible assumption, similar to the temperature-independent treatment of the dielectric constant in coarse-grained simulations; however, the dielectric constant is expected to vary when moving from the membrane interior to the exterior<sup>64</sup> which can be addressed only by all-atom MD simulations. Third, and in continuation from the point above, the presence of large extracellular domains means two or more effective dielectric constants need to be considered, a feature which is not introduced in the model currently. Despite these limitations, the rapidity of the bWSME approach (the total partition function can be calculated in a few minutes), the physical rigor of the model, and the ability to reproduce multiple experimental data in a (semi-)quantitative manner, makes this thermodynamic framework quite appealing. In addition, the folded status of select residues or structural elements can be modulated by the conformational entropy parameter to capture specific or unique experimental observations not evident from the static structure. Alanine-scanning mutagenesis using the bWSME model has the potential to provide mechanistic insights into the extent to which the intramolecular network determines allosteric responses and the role of conformational ensembles in determining the same. Thus, a synergistic use of the bWSME model predictions with experiments can provide a holistic picture of the unique structure–ensemble–function relationship prevalent in GPCRs.

## Methods

### GPCR database

Sixty-seven high-resolution structures were downloaded from the GPCR-EXP database<sup>65</sup> (experimentally solved GPCR structures) out of which only those structures that consisted primarily of the transmembrane domains were selected. In other words, those structures with large intracellular or extracellular domains were discarded as the hydrophilic environments in which such domains exist and the hydrophobic environment within the lipid bilayer cannot be modeled simultaneously using a single dielectric constant (vide infra). Structures with large intracellular or extracellular loops that are not amenable to modeling using the Robetta server<sup>66</sup> were also discarded. The pruning eventually resulted in a database of 45 GPCRs; 41 from humans and one each from bovine, mouse, rat, and viral taxa (Supplementary Table 1). Any missing residues or short loops were again modeled using the Robetta server. The sequences of these missing segments, including the third intracellular loop (ICL3), which is often replaced with a fusion protein to facilitate crystallization, were obtained from UniProtKB. Missing N- and C-terminal segments in structures obtained from truncated GPCR constructs were not modeled. Sequence modifications already present in the original PDB structures, including thermostabilizing mutations, were left unaltered. Of the mammalian GPCRs, 39 belong to class A, the rhodopsin-like receptor family. Classes B, C, and F are represented by the calcitonin receptor, two metabotropic glutamate receptors, and Frizzled-4, respectively. These 45 structures are of the GPCRs in their inactive states, bound only to a ligand (agonist or antagonist) on the extracellular side. Note that we use the term “inactive” to refer to “non-G-protein-bound” structures. Rhodopsin is considered only in the retinal-free form. The database additionally contains the structures of 8 GPCRs in their active conformations, with their transmembrane helices having undergone rearrangements characteristic of the active states

through the binding of both an agonist on the extracellular side and a transducer protein or an antibody on the intracellular side.

### Wako–Saitô–Muñoz–Eaton (WSME) model

The WSME model is a structure-based statistical mechanical model that employs a Gō-like treatment for its energetics, i.e., only those contacts or interactions that are present in the native structure are assumed to influence the folding mechanism, and is therefore entirely native-centric in its description (non-native interactions are not considered). While the model is explained in detail in many works before<sup>39,40,67</sup>, it is briefly discussed here. In the classic WSME treatment<sup>36,37</sup>, every residue is assumed to sample two conformational substates, folded (represented by the binary variable 1) and unfolded (0), resulting in  $2^N$  microstates or conformations for an  $N$ -residue protein. In the current treatment, we employ a computationally less-intensive approach wherein only stretches of consecutive residues, termed blocks, are considered as the folding unit<sup>67</sup>. For example, for a 300-residue protein assuming a block-size of 3 will reduce the number of folding units to 100 blocks, instead of 300 units. Furthermore, the instantaneous ensemble is considered to be constituted from single stretches of folded blocks (single sequence approximation or SSA), two stretches of folded blocks (double sequence approximation or DSA), and DSA allowing for interactions between the folded islands (DSA with loop, or DSAw/L)<sup>67,68</sup>. One can imagine each of these microstates to be an array of strings with 1s and 0s defining the regions that are structured and hence the extent of structure; the constraint is that there can be at most only two islands of ones (DSA or DSAw/L) (Supplementary Fig. 1). The latter DSAw/L approximation is critical as it allows for two folded islands to interact with each other if they do so in the folded structure—the precise stability of such microstates will be determined by the relative balance between stabilizing interactions within and between the folded islands and the entropic cost of fixing residues in the intervening disordered loop. The resulting bWSME model (b standing for block) has been tested on multiple proteins and has provided insights into folding mechanisms and function in an experimentally consistent manner<sup>39,40,67</sup>. In the current work, a fixed block length of 4 (i.e., four consecutive residues and ensuring that the block definition does not span two different secondary structure elements) has been employed for all GPCRs. The total number of microstates within the model approximation is the sum of the binomial coefficients  $C_2^{N_b+1} + 2C_4^{N_b+1}$  where  $N_b$  is the total number of blocks (Supplementary Fig. 1).

The total partition function ( $Z$ ) for the bWSME model is calculated as:

$$Z = \sum_{i=1}^n w_i = \sum_{i=1}^n \exp(-\Delta G_i/RT) \quad (2)$$

where  $n$  is the total number of microstates (i.e., microstates defined by SSA, DSA, and DSAw/L),  $w_i$  is the statistical weight of state  $i$ ,  $R$  is the gas constant, and  $T$  is the temperature. The free energy of every microstate with structure between and involving blocks  $p$  and  $q$  ( $p, q$ ) is

$$\Delta G_{p,q} = \Delta G_{p,q}^{\text{stab}} - T\Delta S_{p,q}^{\text{conf}} \quad (3)$$

The stabilization free energy  $\Delta G_{p,q}^{\text{stab}}$  includes contributions from vdW interactions (interaction energy  $\xi$  for the vdW contacts identified from the ligand-free native structure with a 5 Å cut-off including nearest neighbors), charge–charge interactions at pH 7.0 without a distance cut-off via the Debye–Hückel formalism, and a contacts-scaled implicit heat capacity term ( $\Delta G_{\text{solv}}$ , calculated as the heat capacity change per native contact  $\Delta C_p^{\text{cont}}$  that is fixed to  $-0.36 \text{ J mol}^{-1} \text{ K}^{-1}$  per native contact)<sup>38,39</sup>. Charge–charge interactions are calculated with an effective dielectric constant of 4 to account for the low dielectric environment of the membrane environment<sup>39,64,69</sup>, similar to the approach

employed for studying water-soluble proteins where the solvent is considered as a uniform high dielectric continuum. The unstructured loops connecting the transmembrane helices are partially exposed to the solvent; we account for the disordered nature of the loops by attributing a higher entropic penalty for ordering (see below) thus recapitulating expectations from the structure. The role of disulfide bridges were not explicitly considered and the GPCRs were assumed to be in the reduced form to be amenable for characterization by the bWSME model.

The entropic penalty incurred for fixing all blocks in the folded conformation for the microstate ( $p, q$ ) is given as,

$$\Delta S_{p,q}^{\text{conf}} = \sum_{i=p}^q \sum_{j=L(i)} \Delta S_j^{\text{conf}} \quad (4)$$

$\Delta S_j^{\text{conf}}$  is the entropic penalty for fixing the residue  $j$  in the folded conformation (fixed at  $-10 \text{ J mol}^{-1} \text{ K}^{-1}$  per residue) while  $L(i)$  includes the set of residues within block  $i$ . An excess entropic penalty ( $\Delta\Delta S$ ) of  $-6.1 \text{ J mol}^{-1} \text{ K}^{-1}$  per residue is additionally assigned to residues identified as coil by STRIDE<sup>70</sup> (mostly the loop regions connecting TM helices)<sup>71</sup>. The entropic penalty of fixing a proline residue in the native conformation is considered to be  $0 \text{ J mol}^{-1} \text{ K}^{-1}$ , owing to its limited conformational flexibility. Partial partition functions are calculated by grouping microstates with a specific number of structured blocks from which the one-dimensional free energy profiles are generated. For example, the effective free energy of states with 30 structured blocks is calculated from:

$$\Delta G_{30} = -RT \ln(Z_{30}/Z) \quad (5)$$

where  $Z_{30}$  is sum of statistical weights of states with 30 structured blocks. For comparison between proteins of different lengths ( $N$ ), the fraction of structured blocks calculated by normalizing the number of structured blocks by the maximum number of structured blocks ( $N_b$ ) in the protein, was employed. A similar calculation was employed to construct 2D landscapes for specific combination of blocks structured in the N- and C-terminal halves of the protein. The folding probability of a specific block/residue  $i$  is calculated from:

$$p_i = \sum_k w_k / Z \quad (6)$$

where  $k$  runs over all microstates in which residue  $i$  is folded. From Eq. (6), the stability ( $s$ ) of residue  $i$  in the context of the structure is:

$$\Delta G_{s,i} = -RT \ln(p_i / (1 - p_i)) \quad (7)$$

and the mean stability of a secondary structure (ss) element as:

$$\langle \Delta G_{ss} \rangle = -RT \ln(\langle p_i \rangle / (1 - \langle p_i \rangle)) \quad (8)$$

where the square brackets average over all residues corresponding to a specific secondary structure element. The heat capacity profiles were predicted via the derivative expression:

$$C_p \cong C_v = 2RT \left( \frac{d \ln Z}{dT} \right) + RT^2 \left( \frac{d^2 \ln Z}{dT^2} \right) \quad (9)$$

For the database of GPCRs in Supplementary Table 1, only the vdW interaction energy ( $\xi$ ) was adjusted such that the resulting heat capacity curve has a peak heat capacity at 333 K (the melting temperature,  $T_m$ ). In specific cases, where two heat capacity peaks were present, the  $\xi$  was adjusted such that the trough between the two peaks falls at 333 K. For GPCRs that have had structures in both active and inactive states experimentally determined, different  $\xi$  values were

employed such that the free-energy difference between the folded and unfolded states of a receptor in both its active and inactive states were equal (i.e., under conditions of iso-stability). Note that ligands are not considered in the analysis and only the information from the polypeptide is employed to generate an ensemble of states and their relative statistical weights. The PDB structures employed and the model outputs, including the MATLAB scripts for analyzing them are available at <https://github.com/AthiNaganathan/GPCR-Landscapes>.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The data that support this study are available from the corresponding authors upon reasonable request. High-resolution GPCR structures were downloaded from the GPCR-EXP database. The following are the original Protein Data Bank accession codes: 1U19, 2LNL, 2RHI, 3ODU, 3V2Y, 3VW7, 4BVN, 4DJH, 4DKL, 4N6H, 4OR2, 4PXZ, 4XES, 4XNV, 4XT1, 4YAY, 4Z35, 4ZJ8, 5DHG, 5LWE, 5NDD, 5TZR, 5UEN, SUIW, SUNF, 5VBL, 5VEW, 5ZBQ, 5ZKP, 5ZTY, 6A94, 6BD4, 6CIR, 6D27, 6FFI, 6GPX, 6HLP, 6IGK, 6IIU, 6M9T, 6ME2, 6ME6, 6NIY, 6QZH, 6RZ6, 5WOP, 4LDE, 6H7N, 6B73, 5CIM, 6OS9, 6DO1, 6D9H, 2XOV, 1THQ. All datasets generated during this study are available in the Github repository [<https://github.com/AthiNaganathan/GPCR-Landscapes>]. Source data are provided with this paper.

## Code availability

The data analysis codes and scripts employed in this study used MATLAB 2020a and PyMol. The basic algorithm, code used for generating free energy profiles, and coupling free energy matrices are available at <https://github.com/AthiNaganathan/GPCR-Landscapes>. The same has been published at the Zenodo repository [<https://doi.org/10.5281/zenodo.7426052>]. Any scripts required for analysis are freely available on request by contacting the corresponding author.

## References

- Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharm. Sci.* **33**, 17–27 (2012).
- Schöneberg, T. et al. Mutant G-protein-coupled receptors as a cause of human diseases. *Pharm. Ther.* **104**, 173–206 (2004).
- Thompson, M. D., Hendy, G. N., Percy, M. E., Bichet, D. G. & Cole, D. E. C. G protein-coupled receptor mutations and human genetic disease. *Methods Mol. Biol.* **1175**, 153–187 (2014).
- Latorraca, N. R., Venkatakrisnan, A. J. & Dror, R. O. GPCR dynamics: structures in motion. *Chem. Rev.* **117**, 139–155 (2017).
- Valentin-Hansen, L., Holst, B., Frimurer, T. M. & Schwartz, T. W. PheVI:09 (Phe6.44) as a sliding microswitch in seven-transmembrane (7TM) G protein-coupled receptor activation. *J. Biol. Chem.* **287**, 43516–43526 (2012).
- Katritch, V., Cherezov, V. & Stevens, R. C. Structure-function of the G protein-coupled receptor superfamily. *Ann. Rev. Pharm. Toxicol.* **53**, 531–556 (2013).
- Hoare, S. R. J., Tewson, P. H., Quinn, A. M., Hughes, T. E. & Bridge, L. J. Analyzing kinetic signaling data for G-protein-coupled receptors. *Sci. Rep.* **10**, 12263 (2020).
- Dror, R. O. et al. Activation mechanism of the B2-adrenergic receptor. *Proc. Natl Acad. Sci. USA* **108**, 18684–18689 (2011).
- Tautermann, C. S., Seeliger, D. & Kriegl, J. M. What can we learn from molecular dynamics simulations for GPCR drug design. *Comput. Struct. Biotechnol. J.* **13**, 111–121 (2015).

11. Miao, Y. & McCammon, J. A. G-protein coupled receptors: advances in simulation and drug discovery. *Curr. Opin. Struct. Biol.* **41**, 83–89 (2016).
12. Sengupta, D., Prasanna, X., Mohole, M. & Chattopadhyay, A. Exploring GPCR-lipid interactions by molecular dynamics simulations: excitements, challenges, and the way forward. *J. Phys. Chem. B* **122**, 5727–5737 (2018).
13. Velgy, N., Hedger, G. & Biggin, P. C. GPCRs: what can we learn from molecular dynamics simulations? *Methods Mol. Biol.* **1705**, 133–158 (2018).
14. Zhou, Q. et al. Common activation mechanism of class A GPCRs. *Elife* **8**, e50279 (2019).
15. Hauser, A. S. et al. GPCR activation mechanisms across classes and macro/microscales. *Nat. Struct. Mol. Biol.* **28**, 879–888 (2021).
16. Granier, S., Kim, S., Fung, J. J., Bokoch, M. P. & Parnot, C. FRET-based measurement of GPCR conformational changes. *Methods Mol. Biol.* **552**, 253–268 (2009).
17. Vilardaga, J.-P. Studying ligand efficacy at G protein-coupled receptors using FRET. *Methods Mol. Biol.* **756**, 133–148 (2011).
18. Shi, P. et al. A genetically encoded small-size fluorescent pair reveals allosteric conformational changes of G proteins upon its interaction with GPCRs by fluorescence lifetime based FRET. *Chem. Commun.* **56**, 6941–6944 (2020).
19. Asher, W. B. et al. Single-molecule FRET imaging of GPCR dimers in living cells. *Nat. Methods* **18**, 397–405 (2021).
20. Zhou, Y., Meng, J., Xu, C. & Liu, J. Multiple GPCR functional assays based on resonance energy transfer sensors. *Front. Cell. Dev. Biol.* **9**, 611443 (2021).
21. Li, S., Lee, S. Y. & Chung, K. Y. Conformational analysis of G protein-coupled receptor signaling by hydrogen/deuterium exchange mass spectrometry. *Methods Enzymol.* **557**, 261–278 (2015).
22. Xiao, K., Chung, J. & Wall, A. The power of mass spectrometry in structural characterization of GPCR signaling. *J. Recept. Signal. Transduct. Res.* **35**, 213–219 (2015).
23. Du, Y. et al. Assembly of a GPCR-G protein complex. *Cell* **177**, 1232.e11–1242.e11 (2019).
24. Kim, H. R. et al. Structural mechanism underlying primary and secondary coupling between GPCRs and the Gi/o family. *Nat. Commun.* **11**, 3160 (2020).
25. Martens, C. & Politis, A. A glimpse into the molecular mechanism of integral membrane proteins through hydrogen-deuterium exchange mass spectrometry. *Protein Sci.* **29**, 1285–1301 (2020).
26. Van Eps, N., Caro, L. N., Morizumi, T. & Ernst, O. P. Characterizing rhodopsin signaling by EPR spectroscopy: from structure to dynamics. *Photochem. Photobiol. Sci.* **14**, 1586–1597 (2015).
27. Kaiser, A. & Coin, I. Capturing peptide-GPCR interactions and their dynamics. *Molecules* **25**, 4724 (2020).
28. Lerch, M. T. et al. Viewing rare conformations of the  $\beta(2)$  adrenergic receptor with pressure-resolved DEER spectroscopy. *Proc. Natl Acad. Sci. USA* **117**, 31824–31831 (2020).
29. Elgeti, M. & Hubbell, W. L. DEER analysis of GPCR conformational heterogeneity. *Biomolecules* **11**, 778 (2021).
30. Reichenwallner, J., Liu, B., Balo, A. R., Ou, W.-L. & Ernst, O. P. Electron paramagnetic resonance spectroscopy on G-protein-coupled receptors: adopting strategies from related model systems. *Curr. Opin. Struct. Biol.* **69**, 177–186 (2021).
31. Bostock, M. J., Solt, A. S. & Nietlispach, D. The role of NMR spectroscopy in mapping the conformational landscape of GPCRs. *Curr. Opin. Struct. Biol.* **57**, 145–156 (2019).
32. Casiraghi, M. et al. NMR analysis of GPCR conformational landscapes and dynamics. *Mol. Cell. Endocrinol.* **484**, 69–77 (2019).
33. Frei, J. N. et al. Conformational plasticity of ligand-bound and ternary GPCR complexes studied by (19)F NMR of the  $\beta(1)$ -adrenergic receptor. *Nat. Commun.* **11**, 669 (2020).
34. Park, S. H. & Lee, J. H. Dynamic G protein-coupled receptor signaling probed by solution NMR spectroscopy. *Biochemistry* **59**, 1065–1080 (2020).
35. Waltenspühl, Y., Ehrenmann, J., Klenk, C. & Plückthun, A. Engineering of challenging G protein-coupled receptors for structure determination and biophysical studies. *Molecules* **26**, 1465 (2021).
36. Wako, H. & Saito, N. Statistical mechanical theory of protein conformation. 2. Folding pathway for protein. *J. Phys. Soc. Jpn.* **44**, 1939–1945 (1978).
37. Muñoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl Acad. Sci. USA* **96**, 11311–11316 (1999).
38. Naganathan, A. N. Predictions from an Ising-like statistical mechanical model on the dynamic and thermodynamic effects of protein surface electrostatics. *J. Chem. Theory Comput.* **8**, 4646–4656 (2012).
39. Naganathan, A. N. & Kannan, A. A hierarchy of coupling free energies underlie the thermodynamic and functional architecture of protein structures. *Curr. Res. Struct. Biol.* **3**, 257–267 (2021).
40. Naganathan, A. N., Dani, R., Gopi, S., Aranganathan, A. & Narayan, A. Folding intermediates, heterogeneous native ensembles and protein function. *J. Mol. Biol.* **433**, 167325 (2021).
41. Larkin, M. A. et al. Version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
42. Holm, L. Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).
43. Cornell, W. D. et al. Generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
44. Min, D., Jefferson, R. E., Bowie, J. U. & Yoon, T.-Y. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* **11**, 981–987 (2015).
45. Huysmans, G. H. M., Baldwin, S. A., Brockwell, D. J. & Radford, S. E. The transition state for folding of an outer membrane protein. *Proc. Natl Acad. Sci. USA* **107**, 4099–4104 (2010).
46. Narayan, A., Gopi, S., Lukose, B. & Naganathan, A. N. Electrostatic frustration shapes folding mechanistic differences in paralogous bacterial stress response proteins. *J. Mol. Biol.* **432**, 4830–4839 (2020).
47. Gopi, S., Lukose, B. & Naganathan, A. N. Diverse native ensembles dictate the differential functional responses of nuclear receptor ligand-binding domains. *J. Phys. Chem. B* **125**, 3546–3555 (2021).
48. Naganathan, A. N. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Curr. Opin. Struct. Biol.* **54**, 1–9 (2019).
49. Manglik, A. et al. Structural insights into the dynamic process of B2-adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
50. Freire, E. The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl Acad. Sci. USA* **96**, 10118–10122 (1999).
51. Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl Acad. Sci. USA* **95**, 9903–9908 (1998).
52. Chowdhury, S. & Chanda, B. Deconstructing thermodynamic parameters of a coupled system from site-specific observables. *Proc. Natl Acad. Sci. USA* **107**, 18856–18861 (2010).
53. Venkatakrisnan, A. J. et al. Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
54. Nygaard, R. et al. The dynamic process of  $\beta(2)$ -adrenergic receptor activation. *Cell* **152**, 532–542 (2013).
55. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC, 2022).
56. Weis, W. I. & Kobilka, B. K. The molecular basis of G protein-coupled receptor activation. *Ann. Rev. Biochem.* **87**, 897–919 (2018).

57. Sounier, R. et al. Propagation of conformational changes during  $\mu$ -opioid receptor activation. *Nature* **524**, 375–378 (2015).
  58. Wang, T. & Duan, Y. Ligand entry and exit pathways in the beta2-adrenergic receptor. *J. Mol. Biol.* **392**, 1102–1115 (2009).
  59. Dror, R. O. et al. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **108**, 13118–13123 (2011).
  60. Wang, T. & Duan, Y. Chromophore channeling in the G-protein coupled receptor rhodopsin. *J. Am. Chem. Soc.* **129**, 6970–6971 (2007).
  61. Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).
  62. Chan, H. S., Zhang, Z., Wallin, S. & Liu, Z. Cooperativity, local-nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models. *Ann. Rev. Phys. Chem.* **62**, 301–326 (2011).
  63. Kluber, A., Burt, T. A. & Clementi, C. Size and topology modulate the effects of frustration in protein folding. *Proc. Natl Acad. Sci. USA* **115**, 9234–9239 (2018).
  64. Tanizaki, S. & Feig, M. A generalized born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J. Chem. Phys.* **122**, 124706 (2005).
  65. Chan, W. K. B. & Zhang, Y. Virtual screening of human class-A GPCRs using ligand profiles built on multiple ligand-receptor interactions. *J. Mol. Biol.* **432**, 4872–4890 (2020).
  66. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta Server. *Nucleic Acids Res.* **32**, W526–W531 (2004).
  67. Gopi, S., Aranganathan, A. & Naganathan, A. N. Thermodynamics and folding landscapes of large proteins from a statistical mechanical model. *Curr. Res. Struct. Biol.* **1**, 6–12 (2019).
  68. Henry, E. R. & Eaton, W. A. Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chem. Phys.* **307**, 163–185 (2004).
  69. Tanizaki, S. & Feig, M. Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model. *J. Phys. Chem. B* **110**, 548–556 (2006).
  70. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502 (2004).
  71. Rajasekaran, N., Gopi, S., Narayan, A. & Naganathan, A. N. Quantifying protein disorder through measures of excess conformational entropy. *J. Phys. Chem. B* **120**, 4341–4350 (2016).
- for the grant MTR/2019/000392 to A.N.N. and acknowledge financial support from the Ministry of Education, New Delhi (Sanction No. 11/9/2019-U.3(A)), and the Centre of Excellence in Biochemical Sensing and Imaging Technologies (CenBioSIIm), Indian Institute of Technology Madras. We acknowledge the use of the computing resources at HPCE, IIT Madras. The authors thank Anirudh Ranganathan for comments on the manuscript.

### Author contributions

A.N.N. designed the research study. S.A. generated the database and performed the simulations. S.A. and A.N.N. analyzed the data, interpreted the results, prepared figures, and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-35790-z>.

**Correspondence** and requests for materials should be addressed to Athi N. Naganathan.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

### Acknowledgements

The authors are grateful for the support of the Science and Engineering Research Board (SERB; Department of Science and Technology, India)