

BaNdyT: Bayesian Network modeling of molecular Dynamics Trajectories

*Elizaveta Mukhaleva^{1,2}, Babgen Manookian¹, Hanyu Chen^{1,2}, Ning Ma¹, Wenyuan Wei^{1,2},
Konstancja Urbaniak¹, Grigoriy Gogoshin¹, Supriyo Bhattacharya^{1*}, Nagarajan Vaidehi^{1,2*},
Arei S. Rodin^{1,2*} and Sergio Branciamore^{1,2*}*

¹Department of Computational and Quantitative Medicine, Beckman Research Institute of the
City of Hope, 1218 S 5th Ave, Monrovia, CA 91016

²Irell and Manella Graduate School of Biological Sciences, Beckman Research Institute of the
City of Hope, 1500 E Duarte Road, Duarte, California, USA

Abstract

Bayesian network modeling (BN modeling, or BNM) is an interpretable machine learning method for constructing probabilistic graphical models from the data. In recent years, it has been extensively applied to diverse types of biomedical datasets. Concurrently, our ability to perform long-timescale molecular dynamics (MD) simulations on proteins and other materials has increased exponentially. However, the analysis of MD simulation trajectories has not been data-driven but rather dependent on the user's prior knowledge of the systems, thus limiting the scope and utility of the MD simulations. Recently, we pioneered using BNM for analyzing the MD

trajectories of protein complexes. The resulting BN models yield novel fully data-driven insights into the functional importance of the amino acid residues that modulate proteins' function. In this report, we describe the BaNDyT software package that implements the BNM specifically attuned to the MD simulation trajectories data. We believe that BaNDyT is the first software package to include specialized and advanced features for analyzing MD simulation trajectories using a probabilistic graphical network model. We describe here the software's uses, the methods associated with it, and a comprehensive Python interface to the underlying generalist BNM code. This provides a powerful and versatile mechanism for users to control the workflow. As an application example, we have utilized this methodology and associated software to study how membrane proteins, specifically the G protein-coupled receptors, selectively couple to G proteins. The software can be used for analyzing MD trajectories of any protein as well as polymeric materials.

Keywords:

Molecular dynamics simulation trajectories, Bayesian network modeling, interpretable machine learning, probabilistic graphical models, G protein, G protein-coupled receptors

Corresponding authors: sbhattacharya@coh.org; nvaidehi@coh.org; arodin@coh.org and sbranciamore@coh.org

Introduction

The dynamics of proteins play a critical role in deciphering their function which in turn is important for designing drugs targeting proteins. Molecular dynamics (MD) simulations are a widely used toolbox to simulate and understand the dynamics of proteins and their complexes¹⁻⁵. In combination with experimental data from NMR, DEER, and other spectroscopic methods MD simulations provide a detailed atomistic-level view of the protein dynamics and their function⁶⁻²². There has been exponential growth in the feasibility of long-timescale MD simulations of proteins and protein complexes that provide rich information on their dynamic properties. At the same time, there is a dire need to shift the usual paradigm of structural bioinformatics from studying single structures to analyzing conformational ensembles.

MD simulation trajectories provide an abundance of high-dimensional data, which makes it a perfect candidate for network-based secondary analysis. Currently, there are multiple network-centric approaches to the analysis of MD simulations data. They can be divided into two main categories: (i) physical proximity based – protein is represented as a graph with residues being nodes, and edges indicating physical interaction between residues with their weight proportional to correlation coefficients extracted from MD simulations²³⁻²⁷; or (ii) interaction energy based – protein residues make the network's nodes and edge weights are based on interaction energies between residues, where strong attractive interactions receive higher weights, and weak interactions lead to disconnection²⁸⁻³³. These methods are focused on the residues as variables, and there are currently no network-centric tools that represent residue pairs as network nodes for inferring and quantifying dependencies between inter-residue distances from dynamics data. The existing methodology either simply concentrates on the residues (not residue pairs) or uses dimensionality reduction techniques (primarily principal component analysis, PCA) to analyze

the residue pairs data^{34–36}. The latter approach suffers from the inherently low interpretability of PCA, and from the related issue of its inability to separate induced and direct (non-transitive) dependencies. Moreover, there is currently no method based on probabilistic relationships between residues/residue contact pairs that can infer dynamic dependencies in a more data-driven, unbiased manner, regardless of the residues' spatial proximity. This gap highlights the need for a framework that can go beyond physical proximity and correlations, enabling a deeper understanding of residue-residue relationships within the broader context of protein dynamics. Therefore, we reasoned that Bayesian Network (BN) modeling (BNM) was a natural fit with the MD simulation data, due to the capacity of BNs to infer nonlinear non-transitive dependencies (suggesting directional causality) among different parts of the protein by analyzing dynamical trends. Recently, we developed a workflow to apply BNM to analyze MD simulation trajectories of large protein complexes, namely the G protein-coupled receptors (GPCRs) coupled to trimeric G proteins³⁷. Applying such data-driven interpretable network models led to the identification of the previously unknown important GPCR:G protein interface residue pairs that contribute significantly to G protein coupling strength and to their selectivity/promiscuity³⁷.

Since BNM inherently generalizes to different proteins and other materials where MD simulations are used, in this report we describe the BaNDyT software package that implements the specialized BNM application to MD simulation trajectories, in general. We believe that this software represents the first comprehensive BNM solution for MD simulations, including advanced features not available with other network models. We include a pragmatic description of how to use the software, and how to interpret the results.

Methodology

BaNDyT workflow

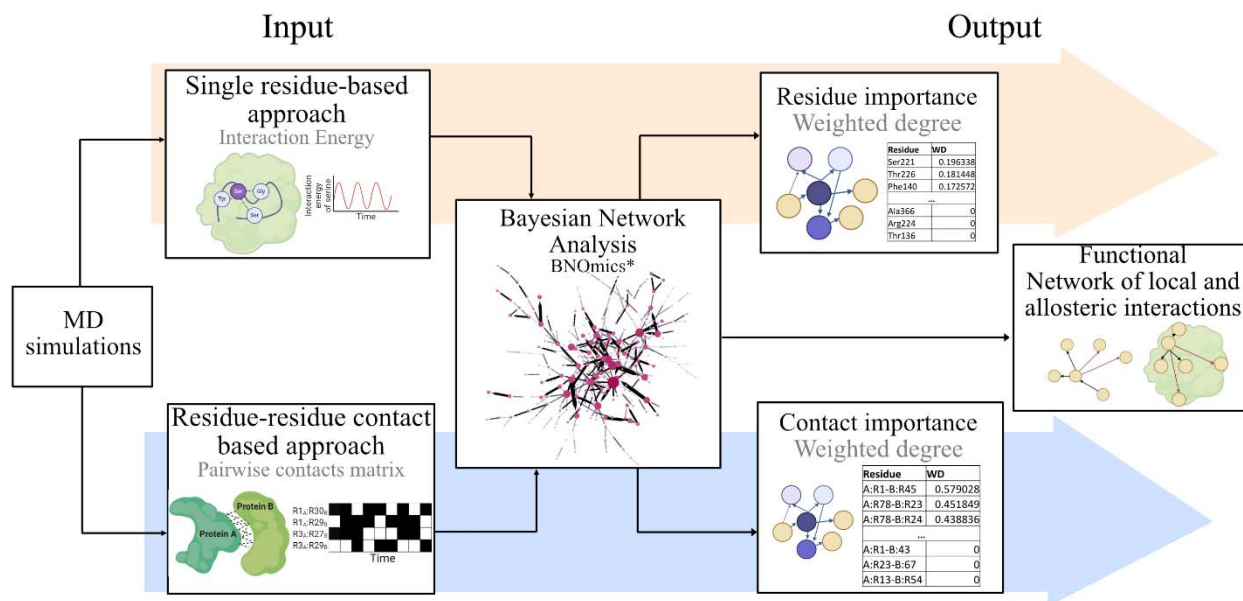


Figure 1. Workflow used in BaNDyT. Step 1 involves performing MD simulations to obtain dynamics trajectories. Step 2 includes extracting data from MD simulation trajectories on single residues and/or pairwise residue-residue interactions as an input for BN analysis. Step 3 uses BNOmics (highly scalable generalist BN modeling software previously developed by us^{38,39}) to construct BN model(s). Step 4 involves visualizing, interpreting, and quantifying the network properties from the BN graph(s) to identify key residues and residue interactions. (BaNDyT can be obtained at GitHub: <https://github.com/bandyt-group/bandyt>)

Step 1: MD Simulations of protein of interest

All-atom molecular dynamics (MD) simulations or any type of enhanced sampling or coarse grain simulations can be performed using standard protocols to generate trajectories that capture the relevant motions and conformational changes^{40–44}. Since the protocols for performing MD simulations vary with the systems simulated, we leave it to the users to formulate their protocols

for performing the simulations. To obtain a stable BN downstream, the user should test for the convergence of the MD simulation trajectories. The users could perform multiple MD runs with different starting velocities to ensure robust statistical sampling of the conformational space that would subsequently yield a stable BN. The number of snapshots stored in each trajectory will affect the network model's robustness – based on our experience, to achieve favorable dimensionality for BNM (given the likely typical protein sizes), we suggest that the user stores snapshots at least every 100ns.

Step 2: Generating input variable values for Bayesian Network Modeling

Properties describing each residue in a protein for every MD simulation snapshot can be used as input variables for deriving BN models. Examples of such residue-based properties include but are not limited to (i) interaction of each residue with the rest of the protein, (ii) packing of each residue with its neighboring residues ⁴⁵, (iii) flexibility of each residue, (iv) torsion angles of each residue, or any other such residue-based attributes.

Properties describing the interaction between pairs of residues in every MD snapshot can also be given as input variables to BNM. This is unique to BNM and has not been done before with other network models. Such pairwise residue properties could be the residue contacts or interaction energies between pairs of residues calculated for each MD snapshot. Parameters such as contact frequency, interaction strength, and changes in contacts over time are analyzed to understand the network of interactions within the protein. Pairwise residue properties can be calculated between residues within a protein or in the interface between proteins forming a complex.

Both types of data can be obtained using standard MD analysis tools readily available^{40–44,46–48}. Such input data combined with BN analysis yields a deeper understanding of the dependencies between inter- or intra-protein interactions.

Step 3: Constructing Bayesian Networks from Molecular Dynamics data with BaNDyT

The data obtained from the two approaches in Step 2 is then used to construct a BN with the BNOmics software, previously developed by us for highly scalable multi-scale BN modeling^{38,39}. Briefly, given a set of input variables, BN modeling generates a DAG (Directed Acyclic Graph) with these variables as nodes connected by the edges that are direct probabilistic dependencies between the variables. BN modeling is an interpretable unsupervised machine learning network-centered methodology, and the residue-based or contact-based properties input variables correspond to the nodes in the network model. Edge strengths indicate probabilistic dependency strengths that can be estimated via a variety of scoring criteria (BNOmics implements the minimum uncertainty (MU) criterion that is based on the resolution limit as well as the customary AIC and MDL/BIC). The sum of all the edge strengths (evaluated on a universal 0-1 scale when using MU³⁸) of a node provides information on its dependencies on adjacent nodes, in the probabilistic space, irrespective of the nodes' structural location. The nodes that are connected to a given node in the BN can be structurally local or non-local (allosteric). Such information on allosteric dependencies (i.e., when the nodes are adjacent in the probabilistic space/BN but remote in the physical space) is valuable in determining the allosteric communication mechanism in proteins^{17,18,49–52}. The sum of edge strengths of each node is also known as “node strength,” or “weighted degree”. It should be re-emphasized here that, in contrast with most other network-centered methods, BN modeling aims to filter out spurious dependencies induced by multicollinearities.

Performing Bayesian Network Modeling of Molecular Dynamic Trajectories data with

BaNdyT

Listing 1 shows the script that is used to run BaNDyT. The following steps are necessary to obtain the BN graph:

```
import sys
sys.path.append('path/to/BaNdyT/folder')
import bandyt

# Read input file
dt = bandyt.read_input_file('input.csv')

# Perform search
srch = bandyt.search(dt)
srch.restarts(nrestarts=50)
srch.dot(path='output')

# Convert to igraph format and get graph properties
bandyt.convert_bn_to_igraph(srch, fout="output.pickle", format="pickle")
bandyt.getGraphProp('output.pickle', 'output')
```

Listing 1. Example of BaNDyT workflow.

Data Loading and Discretization: The input data, derived from MD simulation trajectories, is loaded using the `bandyt.read_input_file` function. In this function, the data containing the continuous variables (for example, residue-centric energy of interaction) is discretized into appropriate bins. By default, BaNDyT utilizes a maximum entropy binning algorithm for discretization⁵³. This approach ensures that each bin contains a roughly equal number of data points, while also maintaining as much randomness as possible in the allocation of data to bins. By maximizing entropy, this method avoids making strong assumptions about the underlying data distribution and ensures that the bins reflect the most uniform possible partition of the data. It is particularly useful in situations where the data distribution is unknown or complex, as it leads to more balanced and informative bins. As the number of observable rare events grows

with arity in the case of the recommended scoring function MU, we set the default number of bins as eight, to capture rare events that are still significant.

BN search (model selection): The discretized data is then used to search for the optimal BN topology using the `bandyt.search` function. There are three scoring functions available in BaNDyT: `mdla` (based on AIC⁵⁴), `mdlb` (based on MDL/BIC⁵⁵), and `mu` (based on MU³⁸). We recommend using MU (`bandyt.mu`) as the scoring function/criterion, due to (i) its superior performance in avoiding false positives (i.e., if the edge, no matter how weak, is in the resulting network, it is highly likely to reflect a real direct dependency) and (ii) edge strengths being estimated on a universal 0-1 scale, which makes downstream comparative analyses possible.

To optimize the number of iterations needed to attain the satisfactory convergence of the BN model selection process, we performed BN reconstruction on proteins of different sizes (Fig. 2A, Table S1). We performed MD simulations on each of these proteins, choosing proteins that vary in sizes from 70 to 1091 residues. We used BaNDyT with MU to generate BN models. We calculated the weighted Hamming distance⁵⁶ between the graphs to assess the improvement in the network topologies after each restart as compared to the original model to evaluate the similarity. For all the eight proteins of varied sizes considered here, the models appear to show significant improvements up to 40 restarts, after which the changes in topology are minimal (average Hamming distance per node is less than 0.01, Fig. 3B). Therefore, to ensure robustness and correct BN model selection, we recommend repeating the search process with at least 50 restarts (Fig. 2A, B; Fig. S1A).

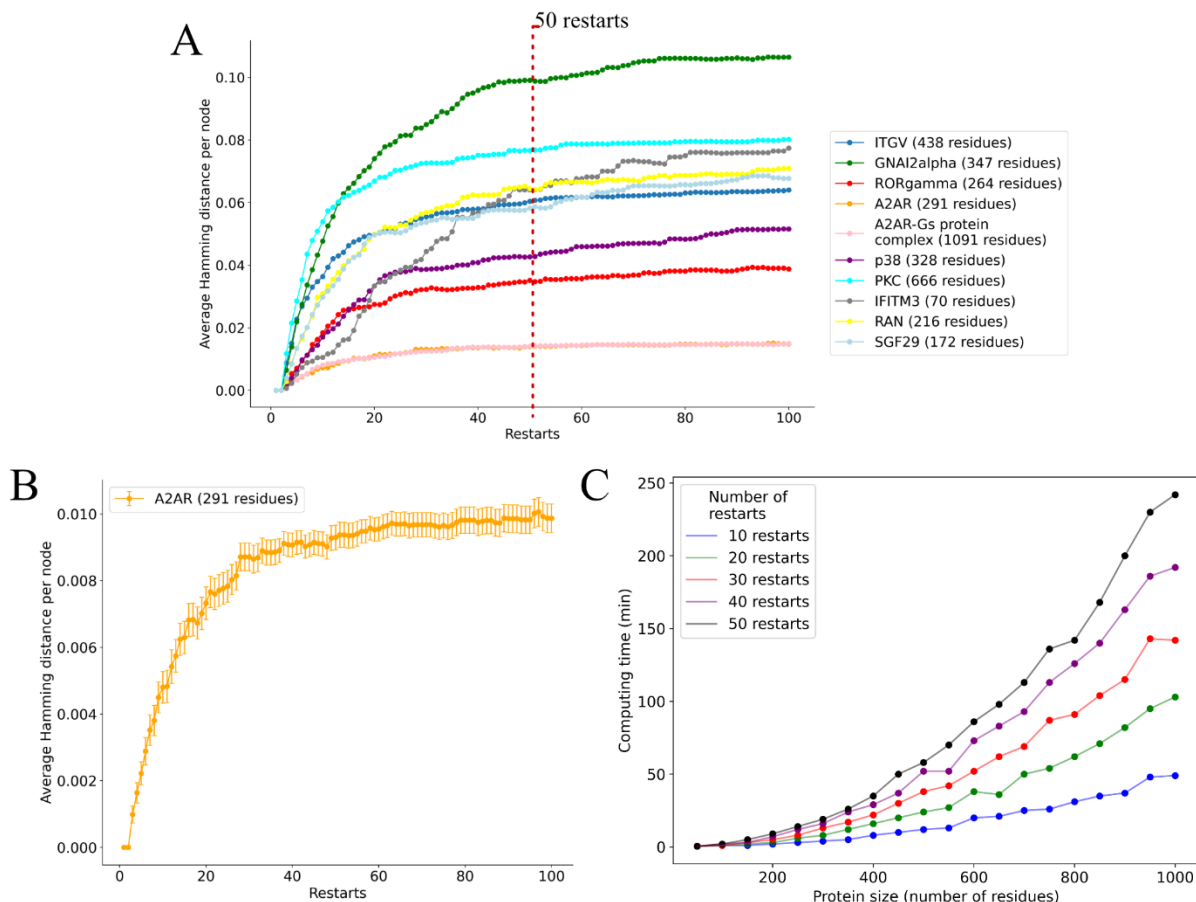


Figure 2. A. Performance and accuracy of Bayesian Network reconstructions for various protein systems. Average Hamming distances per node for BN reconstructions of various proteins with the trajectory length of 5000 frames. Reconstruction for each protein system was replicated 50 times with 100 restarts. The dashed red line indicates the recommended minimal number of restarts. **B.** Average Hamming distances per node for BN reconstructions of A_{2A} receptor. **C.** Computing time of BN models for protein systems of different sizes (on a high-performance cluster with 10G memory and 4 nodes x NVIDIA A100 GPU with 2 x Intel(R) Xeon(R) Platinum 8368 CPU).

Visualizing the BN and Scoring the BN Given the Data: The BN can be visualized using the dot method or using network visualization software such as Cytoscape⁵⁷, which generates a

graphical representation of the network as shown in Fig. 3A. The BN score, obtained with the score_net method, provides a quantitative measure of the network's fit to the data.

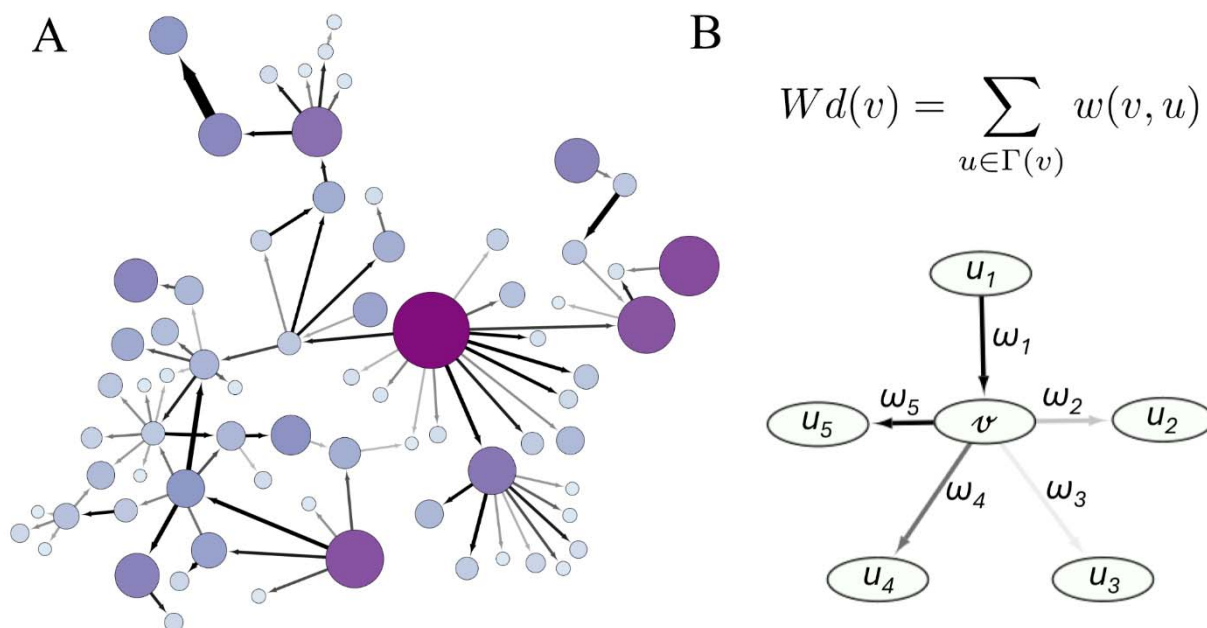


Figure 3. Bayesian Network graph as a main BaNDyT output. **A.** Visual representation of BN graph. The size and color of the node is proportional to the weighted degree of the nodes. The thickness of the arrow is proportionate to the edge weight. **B.** Formula for computing the weighted degree W_d .

The main BaNDyT output is a BN model which can be represented as a DAG, such as shown in Fig.3A. The graph depicts the probabilistic dependencies between the variables (residues or residue contacts) within the protein. Nodes in the graph represent individual variables, while edges represent the direct (non-transitive, non-spurious) dependencies between them. This representation allows us to infer how changes in one region of the protein might influence other local or distant regions, facilitating the identification of key residues and interactions that are critical for the protein function. As a measure of variable importance, local context network

properties associated with nodes (such as betweenness, closeness, etc.), can be extracted. In our case, we have observed that weighted degree shows a strong correlation with experimental data to provide insight into the functional importance of residues and protein-protein interactions³⁷.

The weighted degree of node v (Fig. 3B) can be computed as

$$Wd(v) = \sum_{u \in \Gamma(v)} w(v, u)$$

where $\Gamma(v)$ is the set of nodes in the immediate neighbor of the node v ; $w(v, u)$ denotes the weight of the edge between node v and node u . The weights reflect the strength or significance of the probabilistic dependencies between the nodes, providing insight into the relative importance of these interactions. In the context of the BNM, the weighted degree of the node helps in identifying how central and “influential” the node is within the local network context. A higher weighted degree indicates that the node has stronger or more significant (or numerous) connections with its neighboring nodes, suggesting that it plays a more central role in maintaining the structure and function of the protein.

Step 4: Interpretation of BN Graphs

The final step involves the interpretation of the BN graphs to gain mechanistic insight into the protein dynamics. Here, we aim to translate back from the probabilistic space to the physical space.

- **Residue Importance:** In the single residue approach, the network can be analyzed to determine the significance of individual residues in the overall structure and function. This is done by evaluating metrics such as weighted degree (Fig. 3B).

- **Contact Importance:** Similarly, for the pairwise approach, the importance of residue-residue contacts is assessed based on the BN weighted degree. This involves identifying key interactions that are critical for the stability, cooperativity, and function of protein.
- **Local and Allosteric Network Interactions:** The BN graph provides information (direct node adjacency) on both local interactions (close spatial proximity) and allosteric interactions (distant residues influencing each other). This probabilistic perspective helps in understanding how local changes can propagate through the protein to affect distant sites, shedding light on mechanisms of allosteric regulation.

Example of single residue-based analysis

Here, we will use the histamine H₁ receptor as an example. MD simulation was started from PDB ID 7DFL⁵⁸, which includes the receptor itself and its coupling partner, trimeric Gq protein. We performed multiple runs of all-atom MD simulations in explicit 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer with cholesterol with a total simulation time of 5 μ s (see Methods for more details on MD simulation setup). The main input data for the single residue-based approach is the interaction energy for each protein residue (Fig. 4A, Table S2). To obtain this data, we have used *gmx energy* module from GROMACS2022 version⁵⁹ to calculate the energy of 273 residues of H₁ receptor. By leveraging this input, a BN was constructed to model the probabilistic dependencies among the residues within the receptor, using previously described parameters (8 bins for discretization of input data, 50 restarts for network convergence; Data S1-2). This network effectively captures intricate interactions and allows for a comprehensive analysis of the residue relationships and their potential impact on receptor function. Upon obtaining the BN graph, we calculated the weighted degree to rank the residues

based on their importance within the network (Fig. 4B, C; Table S3). The top 25 percentile of residues, as determined by their weighted degree, are identified as the most important. These key residues were further analyzed to categorize the edges in the network into neighboring interactions and allosteric interactions, based on structural information. Notably, residues within the ligand-binding site were found to have direct connections (in the probabilistic/BN space) with the G protein binding interface and mediator regions, underscoring that minor changes in one location can propagate to distant regions, significantly impacting the receptor's function (Fig. 4B, C).

One of the most important aspects of BN graph analysis is the ability to analyze a community of residues that have established probabilistic relationships. Fig. 4D, E presents an example of the Markov neighborhood (an immediate-adjacency subset/simplification of the Markov blanket) of an important node, residue Q171 in extracellular loop 2. Markov neighborhood of a node refers to the set of all its adjacent nodes (neighbors); the state of the given node is conditionally independent of all other nodes in the graph, given the states of the nodes in its Markov blanket. As seen from the network and structural representation (Fig. 4C, D), the network relationships of this residue can be separated into neighboring and allosteric. Although this residue is not part of the immediate ligand pocket, it evidently influences other functionally important residues, such as 6.36 (G protein binding residue) and 4.56 (ligand binding site residue, mutations that affect protein function⁶⁰). Thus, the BN model provides valuable insights into how specific residues, even the ones that are not located in the functionally important region, can play pivotal roles in the receptor's functionality through their influence on the network of interactions.

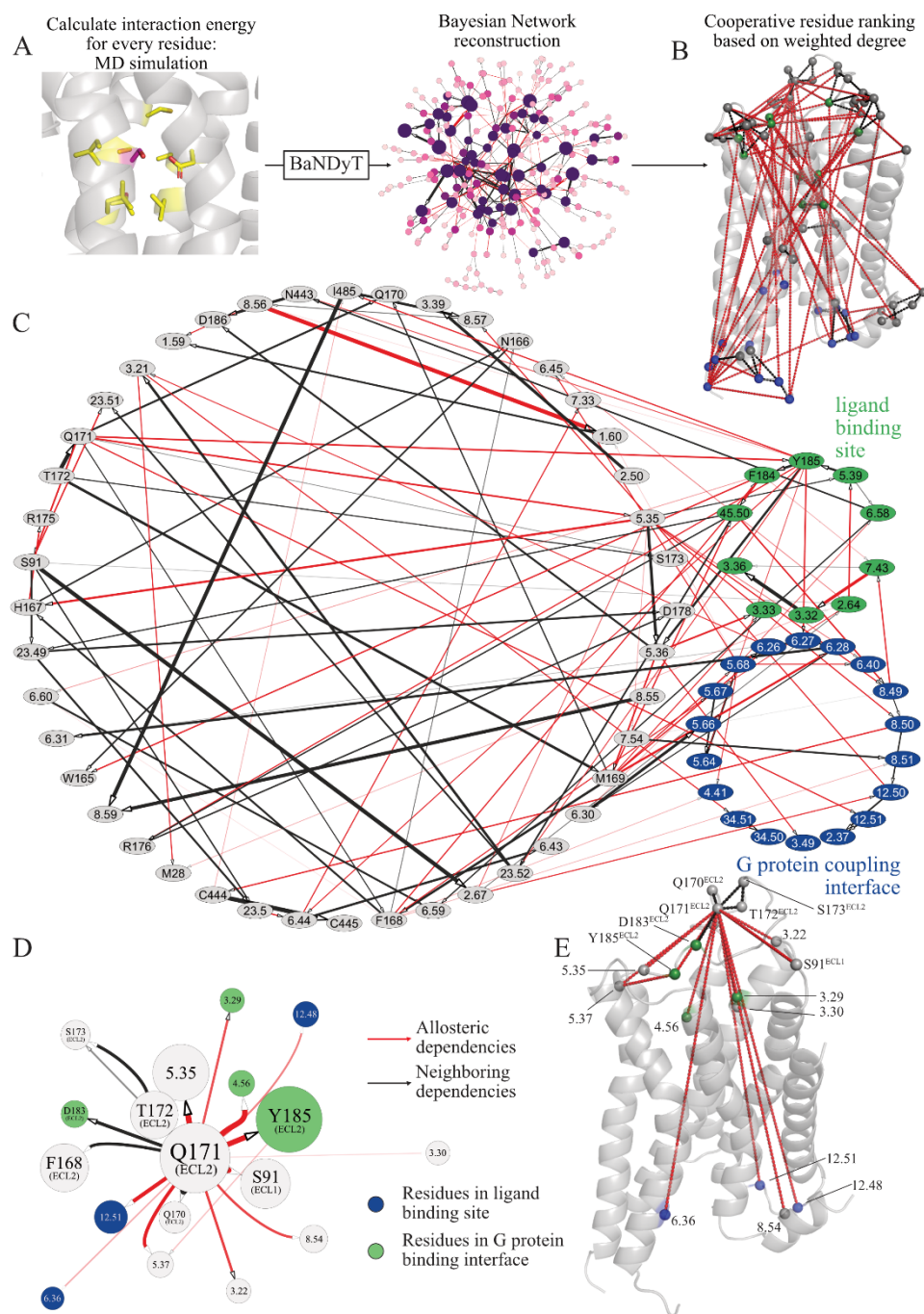


Figure 4. Bayesian Network modeling and structural analysis of H₁ receptor. **A.** BN modeling of H₁ receptor network: interaction energies of each residue in H₁R were used with BaNDyT to obtain the BN graph. **B.** Structural representation of subnetwork of top 25 percentile of nodes based on weighted degree (in C). Ligand-binding site residues are colored green, G

protein binding interface residues - blue. Red dash lines represent allosteric probabilistic relationships, black – neighboring relationships. **C.** BN subnetwork of the most important residues (top 25 percentile) based on the weighted degree. Green nodes correspond to ligand-binding site residues, blue nodes – G protein binding interface residues. Red arrows represent allosteric relationships, black – neighboring relationships. **D.** The Markov neighborhood of residue Q171. Size is proportional to weighted degree; arrow thickness is proportional to edge weight. Red arrows represent allosteric relationships, black – neighboring relationships. Green nodes correspond to ligand-binding site residues, blue nodes – G protein binding interface residues. **E.** Structural representation of the network depicted in D.

Example of pairwise contact-based analysis

Here, we will use the muscarinic receptor 1 (M_1) coupled to the Gq protein as an example. Similarly to H_1 receptor, we have performed all-atom MD simulations of GPCR:G protein complex (PDB ID: 6OIJ⁶¹) in POPC membrane according to the protocol, as described in Methods. To investigate the protein-protein residue contacts within this complex, we utilized GetContacts⁶², which provided a detailed map of these interactions (Fig. 5A). GetContacts-generated output was converted into a binary matrix of residue contacts fingerprints, where 1 represents presence of the contacts in the trajectory frame, and 0 represents absence. Unlike conventional methods, we did not apply any frequency cutoffs, as BN analysis is capable of capturing significant rare events (infrequent contacts) from transient interactions (recall that the MU criterion is optimized for guarding against false positives). This resulted in the dataset of 427 residue-residue contacts with 25,000 trajectory frames as an input for BaNDyT (Table S4). BaNDyT was run using default parameters of 50 restarts, while discretization was not needed

due to data already being binary. The primary outcome of our BN analysis was the BN graph, illustrating the direct dependencies between variables—in this case, the residue contacts (Fig. 5B; Data S3-4). The weighted degree of each node was used as a measure of importance within the network, reflecting the cooperativity of protein-protein interactions observed throughout the MD simulation trajectory (Table S5).

We identified the top 25 percentile of nodes with the highest weighted degrees as the most cooperative interactions, which were subsequently analyzed in the context of protein biology (Fig. 5B, Table S5). This approach has been previously applied to the analysis of six different GPCR protein complexes, revealing significant differences in how various G protein coupling types contribute to the cooperativity of these interactions and highlighting the prevalence of certain types of contacts across different coupling types³⁷.

Recall that the BN graph connects variables in the probabilistic space without considering their structural positions, meaning that contacts that are not in close physical proximity can still be directly connected within the network. This means that even distant contacts may influence each other during the MD trajectory. To further investigate these interactions, we categorized BN dependencies into two groups: neighboring dependencies, where the C α atoms of the GPCR and G α subunit residues in two contacts are within 10Å of each other, and allosteric dependencies, where these atoms are more than 10Å apart (Fig. 5C).

Both allosteric and neighboring dependencies contribute to the importance of a node (contact) within the network. For example, in Fig. 5D, E, we explore the Markov neighborhood of the most cooperative contact in the M₁ protein complex, 5.70_G.H4.20. While this contact is positioned in the core region of the G protein, it has multiple interactions with the contacts formed by H5 helix in the network graph (Fig. 5E). Given the established importance of H5 helix

interactions in G protein coupling, disruptions to contact 5.70_G.H4.20 may have profound effects on the dynamics of dependent H5-helix contacts, potentially compromising GPCR-G protein interaction. Our prior research highlights the core's heightened influence in Gs proteins over Gq proteins³⁷. As the M₁ complex comprises a Gq protein chimera with a Gs protein core⁶¹, this configuration could manifest Gs-like characteristics in the interplay between the core and the H5 tip.

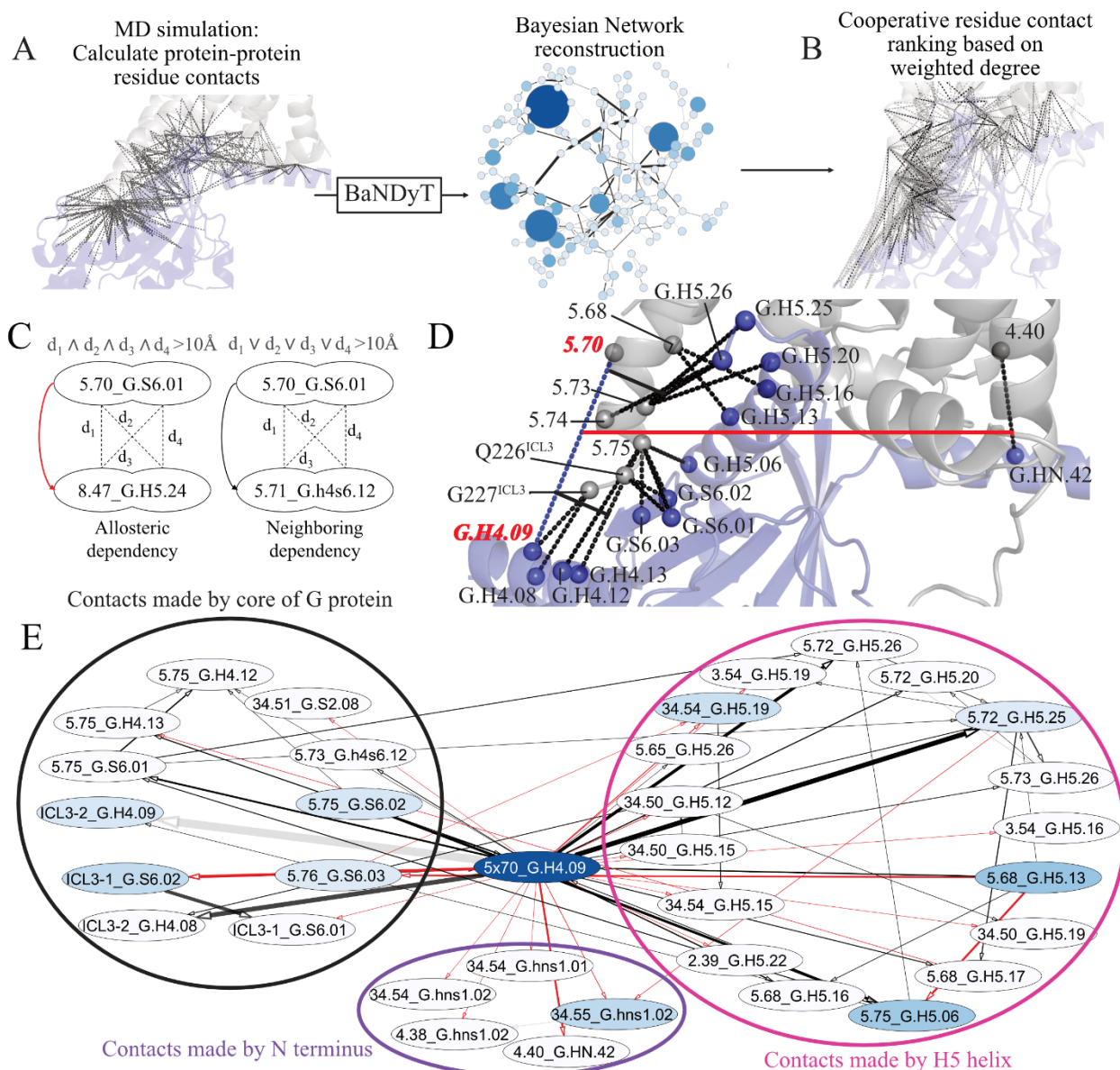


Figure 5. Bayesian Network modeling and structural analysis of M₁:Gq protein residue

contact network. A. BN modeling of M₁:Gq protein contact network: contact fingerprints were used with BaNDyT to obtain the BN graph. **B.** BN graph of M₁:Gq protein residue contact network. Size and color intensity are proportional to the weighted degree of the node, thickness of the arrows is proportional to the edge strength. **C.** Schematic representation of neighboring and allosteric dependencies. Nodes (GPCR:G protein residue contacts) are represented by oval shapes. The pairwise distance between nodes is marked as a black dashed line. A node dependency is considered as neighboring if at least one of the distances (d_1, d_2, d_3, d_4) is less than 10Å and allosteric if all distances (d_1 to d_4) are larger than 10Å. **D.** Structural representation of nodes (contacts) that are connected to node 5.70_G.H4.20 in the BN graph with edge weight higher than 0.01. M₁ is represented as grey cartoon, Gq protein – as blue cartoon. C α atoms of residues involved in the contacts are shown in spheres, black dash lines represent contacts. Red solid line shows allosteric dependency, black solid line – neighboring. **E.** The Markov neighborhood of contact 5.70_G.H4.09. The nodes are grouped by their location in the G protein structure: N terminus, core, or H5 helix. Node color is proportional to weighted degree; arrow thickness is proportional to edge strength. Red arrows represent allosteric relationships, black – neighboring relationships.

Concluding remarks

Here, we have presented BaNDyT, a novel Bayesian Network Modeling software specifically designed for the analysis of MD simulation trajectories. By leveraging the power of BNM, BaNDyT provides an intrinsically interpretable, scalable and fully data-driven unsupervised machine learning approach to uncovering functional relationships between residues and residue pairs in proteins, going beyond traditional MD analysis methods. Our approach addresses several

limitations in the current network-based methodologies, such as their reliance on residue-centric variables and the surfeit of spurious dependencies. With the ability to capture both local and allosteric dependencies, BaNDyT provides a comprehensive framework for analyzing the dynamic interplay within proteins and their complexes, as demonstrated through its application to GPCRs and their selective coupling to G proteins. Furthermore, BaNDyT's versatile Python interface makes it a robust tool for the broader scientific community to explore MD simulation data across various systems, including polymeric materials and other complex biomolecular structures. The resulting Bayesian networks offer new insights into key residues and interactions that are critical for protein stability, function, and allosteric regulation, paving the way for more targeted and effective therapeutic interventions.

Methods

Preparation of Protein Structures for MD simulations

All protein and protein complexes were prepared for MD simulations using the corresponding experimental structures: A_{2A}R:Gs protein complex – 6GDG⁶³, H₁R-Gq protein complex – 7DFL⁵⁸, M₁R-Gq protein complex - 6OIJ⁶¹, SGF29 – 3ME9⁶⁴, ITGV – 3IJE⁶⁵, RORgamma – 4WLB⁶⁶, PKC – 3IW4⁶⁷, GNAI2 - 6CRK⁶⁸, RAN - 2MMC⁶⁹. Mutations in structures were reverted to their wild-type residues using Maestro (Schrödinger Release 2020-1: Maestro, Schrödinger, LLC, New York, 2020). Ligands were parameterized using ParaChem (<https://cgenff.umaryland.edu>). Missing side chains and loops (fewer than 5 residues absent) were modeled into the 3D structures proteins. Residues within 5 Å of mutation sites were minimized using MacroModel, with position restraints applied to all backbone atoms. Protein termini were capped with neutral acetyl and methylamide groups, and histidine protonation states were assigned via the Maestro protein preparation wizard. In case of A_{2A}R and H₁, receptors

were embedded into an explicit POPC bilayer membrane using the PPM 2.0 function from the Orientation of Proteins in Membranes (OPM) tool ⁷⁰, and hydrated with TIP3P water containing 0.15 M NaCl using CHARMM-GUI ^{71,72}. The final system dimensions were approximately 120 Å × 120 Å × 170 Å, and all systems were characterized using the CHARMM36m force field ⁷³.

MD Simulations for Proteins and Protein complexes

All molecular dynamics simulations were conducted using the GROMACS 2019 ⁴⁸ package with a 2 fs integration timestep. The prepared systems underwent energy minimization with position restraints of 10 kcal/mol·Å² applied to the heavy atoms of the proteins, ligand, and lipids if present. A 1-ns heating phase followed, gradually raising the temperature from 0 K to 310 K under the NVT ensemble with the Nosé-Hoover thermostat. This was followed by equilibration in the NPT ensemble, where the initial 1-ns run retained the 10 kcal/mol·Å² restraints, which were then reduced incrementally (from 10 to 5, and finally to 1 kcal/mol·Å²) in 5-ns steps. The final equilibration phase was a 50-ns simulation without position restraints. The last snapshot of this equilibration served as the starting point for five production runs, each 1 μs in length, initiated with randomly generated velocities. The pressure was controlled at 1 bar using the Parrinello-Rahman method ⁷⁴. Nonbonded interactions were calculated with a 12 Å cutoff, and long-range interactions were handled using the Particle Mesh Ewald (PME) method ⁷⁵. The LINCS algorithm constrained all bonds and angles of water molecules.

Calculation of Residue Interaction Energy

Ensemble trajectories were used to analyze interaction energies between protein residues. Interaction energies for each residue with the rest of the protein were calculated using the GROMACS "energy" tool ^{48,59}. The total nonbonded interaction energy, comprising short-range

(within 12 Å) coulombic and van der Waals forces, was extracted from the energy log file and summed to obtain the overall nonbonded interaction energy.

Calculating Fingerprints of Pairwise Interactions Between M₁R and Gαq Protein

Pairwise residue contacts between M₁R and Gq protein were analyzed using the Python script library "GetContacts" ⁶² (<https://www.github.com/getcontacts>). This tool was employed to identify different types of interactions, including salt bridges (cutoff < 4.0 Å), hydrogen bonds (cutoff < 3.5 Å with an angle < 70°), van der Waals contacts (difference < 2 Å), π-stacking interactions (distance < 7.0 Å with an angle < 30° between aromatic planes), and cation-π interactions (distance < 6.0 Å with an angle < 60°). The analysis was conducted over the 5μs trajectory, excluding lipids, water and ions. The atom selection groups were aligned with the relevant residues from both the GPCR and the G protein α-subunit. Each residue was mapped to its corresponding generic residue number using BW numbering for the GPCR and Common G protein numbering for G proteins. Transmembrane helix ends were adjusted using BW numbering. Custom Python scripts were used for one-hot encoding, generating binary contact fingerprints for each simulation frame, where "1" indicates a contact and "0" represents its absence.

Supplementary Material

A Python-based Jupyter Notebook that contains a tutorial for BaNDyT is provided in the supplementary material and can be accessed at <https://github.com/bandyt-group/bandyt-tutorial>.

Data and Software Availability

BaNDyT is available on GitHub at: <https://github.com/bandyt-group/bandyt>. A demo with instructions on how to use the software is available on GitHub at: <https://github.com/bandyt-group/bandyt-tutorial>. An XLSX file containing interaction energy data and a CSV file with a

residue-residue contact matrix, both derived from molecular dynamics simulations and used as inputs for the BaNDyT analysis, are provided in the Supplementary Information (SI). Outputs of BaNDyT for showcases are provided as CSV (network properties), GRAPHML and PICKLE files (graphs) in the SI. Full simulation data will be provided by the authors upon request.

References

- (1) Karplus, M.; McCammon, J. A. Protein Structural Fluctuations during a Period of 100 Ps. *Nature* **1979**, 277 (5697), 578–578. <https://doi.org/10.1038/277578a0>.
- (2) Singharoy, A.; Maffeo, C.; Delgado-Magnero, K. H.; Swainsbury, D. J. K.; Sener, M.; Kleinekathöfer, U.; Vant, J. W.; Nguyen, J.; Hitchcock, A.; Isralewitz, B.; Teo, I.; Chandler, D. E.; Stone, J. E.; Phillips, J. C.; Pogorelov, T. V.; Mallus, M. I.; Chipot, C.; Luthey-Schulten, Z.; Tieleman, D. P.; Hunter, C. N.; Tajkhorshid, E.; Aksimentiev, A.; Schulten, K. Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell* **2019**, 179 (5), 1098-1111.e23. <https://doi.org/10.1016/j.cell.2019.10.021>.
- (3) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. Molecular Dynamics Simulations of Large Macromolecular Complexes. *Curr Opin Struct Biol* **2015**, 31, 64–74. <https://doi.org/10.1016/j.sbi.2015.03.007>.
- (4) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, 1850 (5), 872–877. <https://doi.org/10.1016/j.bbagen.2014.10.019>.
- (5) Cao, F.; von Bülow, S.; Tesei, G.; Lindorff-Larsen, K. A Coarse-grained Model for Disordered and Multi-domain Proteins. *Protein Science* **2024**, 33 (11). <https://doi.org/10.1002/pro.5172>.

- (6) Gibbs, E. B.; Cook, E. C.; Showalter, S. A. Application of NMR to Studies of Intrinsically Disordered Proteins. *Arch Biochem Biophys* **2017**, *628*, 57–70. <https://doi.org/10.1016/j.abb.2017.05.008>.
- (7) Jiang, Y.; Kalodimos, C. G. NMR Studies of Large Proteins. *J Mol Biol* **2017**, *429* (17), 2667–2676. <https://doi.org/10.1016/j.jmb.2017.07.007>.
- (8) Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M. NMR-Based Methods for Protein Analysis. *Anal Chem* **2021**, *93* (4), 1866–1879. <https://doi.org/10.1021/acs.analchem.0c03830>.
- (9) Jiang, L.; Tian, Y.; Zhang, H.; Liu, S. Molecular-Level Insight into the Effects of Low Moisture and Trehalose on the Thermostability of β -Glucosidase. *Food Chem* **2024**, *460*, 140607. <https://doi.org/10.1016/j.foodchem.2024.140607>.
- (10) Rejnowicz, E.; Batchelor, M.; Leen, E.; Ahangar, M. S.; Burgess, S. G.; Richards, M. W.; Kalverda, A. P.; Bayliss, R. Exploring the Dynamics and Interactions of the N-Myc Transactivation Domain through Solution Nuclear Magnetic Resonance Spectroscopy. *Biochemical Journal* **2024**, *481* (21), 1535–1556. <https://doi.org/10.1042/BCJ20240248>.
- (11) Zhang, X.; Tamaki, H.; Kikukawa, T.; Fujiwara, T.; Matsuki, Y. Structural Changes of Natronomonas Pharaonis Halorhodopsin in Its Late Photocycle Revealed by Solid-State NMR Spectroscopy. *Biophys Chem* **2024**, *315*, 107329. <https://doi.org/10.1016/j.bpc.2024.107329>.

(12) Zhao, J.; Elgeti, M.; O'Brien, E. S.; Sár, C. P.; Daibani, A. E.; Heng, J.; Sun, X.; Che, T.; Hubbell, W. L.; Kobilka, B. K.; Chen, C. Conformational Dynamics of the μ -Opioid Receptor Determine Ligand Intrinsic Efficacy. April 29, 2023. <https://doi.org/10.1101/2023.04.28.538657>.

(13) Tang, Q.; Sinclair, M.; Hasdemir, H. S.; Stein, R. A.; Karakas, E.; Tajkhorshid, E.; Mchaourab, H. S. Asymmetric Conformations and Lipid Interactions Shape the ATP-Coupled Cycle of a Heterodimeric ABC Transporter. *Nat Commun* **2023**, *14* (1), 7184. <https://doi.org/10.1038/s41467-023-42937-5>.

(14) Bizet, M.; Byrne, D.; Biaso, F.; Gerbaud, G.; Etienne, E.; Briola, G.; Guigliarelli, B.; Urban, P.; Dorlet, P.; Kalai, T.; Truan, G.; Martinho, M. Structural Insights into the Semiquinone Form of Human Cytochrome P450 Reductase by DEER Distance Measurements between a Native Flavin and a Spin Labelled Non-canonical Amino Acid. *Chemistry – A European Journal* **2024**, *30* (19). <https://doi.org/10.1002/chem.202304307>.

(15) Tani, K.; Maki-Yonekura, S.; Kanno, R.; Negami, T.; Hamaguchi, T.; Hall, M.; Mizoguchi, A.; Humbel, B. M.; Terada, T.; Yonekura, K.; Doi, T. Structure of Endothelin ETB Receptor–Gi Complex in a Conformation Stabilized by Unique NPxxL Motif. *Commun Biol* **2024**, *7* (1), 1303. <https://doi.org/10.1038/s42003-024-06905-z>.

(16) Fatoki, T. H.; Balogun, T. C.; Ojewuyi, A. E.; Omole, A. C.; Olukayode, O. V.; Adewumi, A. P.; Umesi, A. J.; Ijeoma, N. P.; Apooyin, A. E.; Chinedu, C. P.; Idowu, I. E.; Isah, M. J. In Silico Molecular Targets, Docking, Dynamics Simulation and Physiologically Based Pharmacokinetics Modeling of Oritavancin. *BMC Pharmacol Toxicol* **2024**, *25* (1), 79. <https://doi.org/10.1186/s40360-024-00804-z>.

(17) Cao, Y.; van der Velden, W. J. C.; Namkung, Y.; Nivedha, A. K.; Cho, A.; Sedki, D.; Holleran, B.; Lee, N.; Leduc, R.; Muk, S.; Le, K.; Bhattacharya, S.; Vaidehi, N.; Laporte, S. A. Unraveling Allostery within the Angiotensin II Type 1 Receptor for $G\alpha_q$ and β -Arrestin Coupling. *Sci Signal* **2023**, *16* (797). <https://doi.org/10.1126/scisignal.adf2173>.

(18) Bhattacharya, S.; Vaidehi, N. Differences in Allosteric Communication Pipelines in the Inactive and Active States of a GPCR. *Biophys J* **2014**, *107* (2), 422–434. <https://doi.org/10.1016/j.bpj.2014.06.015>.

(19) Vaidehi, N.; Bhattacharya, S.; Larsen, A. B. Structure and Dynamics of G-Protein Coupled Receptors; 2014; pp 37–54. https://doi.org/10.1007/978-94-007-7423-0_3.

(20) Vaidehi, N. Dynamics and Flexibility of G-Protein-Coupled Receptor Conformations and Their Relevance to Drug Design. *Drug Discov Today* **2010**, *15* (21–22), 951–957. <https://doi.org/10.1016/j.drudis.2010.08.018>.

(21) Nedjai, B.; Li, H.; Stroke, I. L.; Wise, E. L.; Webb, M. L.; Merritt, J. R.; Henderson, I.; Klon, A. E.; Cole, A. G.; Horuk, R.; Vaidehi, N.; Pease, J. E. Small Molecule Chemokine Mimetics Suggest a Molecular Basis for the Observation That CXCL10 and CXCL11 Are Allosteric Ligands of CXCR3. *Br J Pharmacol* **2012**, *166* (3), 912–923. <https://doi.org/10.1111/j.1476-5381.2011.01660.x>.

(22) Sandhu, M.; Cho, A.; Ma, N.; Mukhaleva, E.; Namkung, Y.; Lee, S.; Ghosh, S.; Lee, J. H.; Gloriam, D. E.; Laporte, S. A.; Babu, M. M.; Vaidehi, N. Dynamic Spatiotemporal Determinants Modulate GPCR:G Protein Coupling Selectivity and

Promiscuity. *Nat Commun* **2022**, *13* (1), 7428. <https://doi.org/10.1038/s41467-022-34055-5>.

(23) Rivalta, I.; Sultan, M. M.; Lee, N.-S.; Manley, G. A.; Loria, J. P.; Batista, V. S. Allosteric Pathways in Imidazole Glycerol Phosphate Synthase. *Proceedings of the National Academy of Sciences* **2012**, *109* (22). <https://doi.org/10.1073/pnas.1120536109>.

(24) VanWart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. Exploring Residue Component Contributions to Dynamical Network Models of Allostery. *J Chem Theory Comput* **2012**, *8* (8), 2949–2961. <https://doi.org/10.1021/ct300377a>.

(25) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R. L.; McCammon, J. A. Allosteric Networks in Thrombin Distinguish Procoagulant vs. Anticoagulant Activities. *Proceedings of the National Academy of Sciences* **2012**, *109* (52), 21216–21222. <https://doi.org/10.1073/pnas.1218414109>.

(26) Ghosh, A.; Vishveshwara, S. A Study of Communication Pathways in Methionyl-TRNA Synthetase by Molecular Dynamics Simulations and Structure Network Analysis. *Proceedings of the National Academy of Sciences* **2007**, *104* (40), 15711–15716. <https://doi.org/10.1073/pnas.0704459104>.

(27) Melo, M. C. R.; Bernardi, R. C.; de la Fuente-Nunez, C.; Luthey-Schulten, Z. Generalized Correlation-Based Dynamical Network Analysis: A New High-Performance Approach for Identifying Allosteric Communications in Molecular Dynamics Trajectories. *J Chem Phys* **2020**, *153* (13). <https://doi.org/10.1063/5.0018980>.

(28) Bhattacharyya, M.; Bhat, C. R.; Vishveshwara, S. An Automated Approach to Network Features of Protein Structure Ensembles. *Protein Science* **2013**, *22* (10), 1399–1416. <https://doi.org/10.1002/pro.2333>.

(29) Serçinoğlu, O.; Ozbek, P. GRINN: A Tool for Calculation of Residue Interaction Energies and Protein Energy Network Analysis of Molecular Dynamics Simulations. *Nucleic Acids Res* **2018**, *46* (W1), W554–W562. <https://doi.org/10.1093/nar/gky381>.

(30) Tiberti, M.; Invernizzi, G.; Lambrughi, M.; Inbar, Y.; Schreiber, G.; Papaleo, E. PyInteraph: A Framework for the Analysis of Interaction Networks in Structural Ensembles of Proteins. *J Chem Inf Model* **2014**, *54* (5), 1537–1551. <https://doi.org/10.1021/ci400639r>.

(31) Chakrabarty, B.; Naganathan, V.; Garg, K.; Agarwal, Y.; Parekh, N. NAPS Update: Network Analysis of Molecular Dynamics Data and Protein–Nucleic Acid Complexes. *Nucleic Acids Res* **2019**, *47* (W1), W462–W470. <https://doi.org/10.1093/nar/gkz399>.

(32) Piovesan, D.; Minervini, G.; Tosatto, S. C. E. The RING 2.0 Web Server for High Quality Residue Interaction Networks. *Nucleic Acids Res* **2016**, *44* (W1), W367–W374. <https://doi.org/10.1093/nar/gkw315>.

(33) Ribeiro, A. A. S. T.; Ortiz, V. MDN: A Web Portal for Network Analysis of Molecular Dynamics Simulations. *Biophys J* **2015**, *109* (6), 1110–1116. <https://doi.org/10.1016/j.bpj.2015.06.013>.

(34) Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D. Dynamical Network of Residue-Residue Contacts Reveals Coupled Allosteric Effects in Recognition, Catalysis, and Mutation. *Proc Natl Acad Sci U S A* **2016**, *113* (17), 4735–4740. <https://doi.org/10.1073/pnas.1523573113>.

(35) Johnson, Q. R.; Lindsay, R. J.; Nellas, R. B.; Fernandez, E. J.; Shen, T. Mapping Allostery through Computational Glycine Scanning and Correlation Analysis of Residue-Residue Contacts. *Biochemistry* **2015**, *54* (7), 1534–1541. <https://doi.org/10.1021/bi501152d>.

(36) Johnson, Q. R.; Lindsay, R. J.; Shen, T. CAMERRA: An Analysis Tool for the Computation of Conformational Dynamics by Evaluating Residue-Residue Associations. *Journal of computational chemistry*. NLM (Medline) July 30, 2018, pp 1568–1578. <https://doi.org/10.1002/jcc.25192>.

(37) Mukhaleva, E.; Ma, N.; van der Velden, W. J. C.; Gogoshin, G.; Branciamore, S.; Bhattacharya, S.; Rodin, A. S.; Vaidehi, N. Bayesian Network Models Identify Cooperative GPCR:G Protein Interactions That Contribute to G Protein Coupling. *Journal of Biological Chemistry* **2024**, *300* (6), 107362. <https://doi.org/10.1016/j.jbc.2024.107362>.

(38) Gogoshin, G.; Rodin, A. Minimum Statistical Uncertainty as Bayesian Network Model Selection Principle. September 20, 2024. <https://doi.org/10.20944/preprints202202.0254.v3>.

(39) Gogoshin, G.; Boerwinkle, E.; Rodin, A. S. New Algorithm and Software (BNOmics) for Inferring and Visualizing Bayesian Networks from Heterogeneous Big

Biological and Genetic Data. *Journal of Computational Biology* **2017**, *24* (4), 340–356.

<https://doi.org/10.1089/cmb.2016.0100>.

(40) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J Comput Chem* **2005**, *26* (16), 1781–1802. <https://doi.org/10.1002/jcc.20289>.

(41) Kunz, A. E.; Allison, J. R.; Geerke, D. P.; Horta, B. A. C.; Hünenberger, P. H.; Riniker, S.; Schmid, N.; van Gunsteren, W. F. New Functionalities in the GROMOS Biomolecular Simulation Software. *J Comput Chem* **2012**, *33* (3), 340–353. <https://doi.org/10.1002/jcc.21954>.

(42) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>.

(43) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoseck, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **2009**, *30* (10), 1545–1614. <https://doi.org/10.1002/jcc.21287>.

(44) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J Comput Chem* **2005**, *26* (16), 1668–1688. <https://doi.org/10.1002/jcc.20290>.

(45) Zhou, Q.; Yang, D.; Wu, M.; Guo, Y.; Guo, W.; Zhong, L.; Cai, X.; Dai, A.; Jang, W.; Shakhnovich, E. I.; Liu, Z.-J.; Stevens, R. C.; Lambert, N. A.; Babu, M. M.; Wang, M.-W.; Zhao, S. Common Activation Mechanism of Class A GPCRs. *Elife* **2019**, *8*. <https://doi.org/10.7554/eLife.50279>.

(46) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations; 2016; pp 98–105. <https://doi.org/10.25080/Majora-629e541a-00e>.

(47) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **2011**, *32* (10), 2319–2327. <https://doi.org/10.1002/jcc.21787>.

(48) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.

(49) Astore, M. A.; Pradhan, A. S.; Thiede, E. H.; Hanson, S. M. Protein Dynamics Underlying Allosteric Regulation. *Curr Opin Struct Biol* **2024**, *84*, 102768. <https://doi.org/10.1016/j.sbi.2023.102768>.

(50) Xiao, S.; Verkhivker, G. M.; Tao, P. Machine Learning and Protein Allostery. *Trends Biochem Sci* **2023**, *48* (4), 375–390. <https://doi.org/10.1016/j.tibs.2022.12.001>.

(51) Wu, N.; Barahona, M.; Yaliraki, S. N. Allosteric Communication and Signal Transduction in Proteins. *Curr Opin Struct Biol* **2024**, *84*, 102737. <https://doi.org/10.1016/j.sbi.2023.102737>.

(52) Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem Rev* **2016**, *116* (11), 6503–6515. <https://doi.org/10.1021/acs.chemrev.5b00590>.

(53) Rosenblueth, E.; Karmeshu; Han Ping Hong. Maximum Entropy and Discretization of Probability Distributions. *Probabilistic Engineering Mechanics* **1987**, *2* (2), 58–63. [https://doi.org/10.1016/0266-8920\(87\)90016-6](https://doi.org/10.1016/0266-8920(87)90016-6).

(54) Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans Automat Contr* **1974**, *19* (6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>.

(55) Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics* **1978**, *6* (2). <https://doi.org/10.1214/aos/1176344136>.

(56) Hamming, R. W. Error Detecting and Error Correcting Codes. *Bell System Technical Journal* **1950**, *29* (2), 147–160. <https://doi.org/10.1002/j.1538-7305.1950.tb00463.x>.

(57) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. <https://doi.org/10.1101/gr.1239303>.

(58) Xia, R.; Wang, N.; Xu, Z.; Lu, Y.; Song, J.; Zhang, A.; Guo, C.; He, Y. Cryo-EM Structure of the Human Histamine H1 Receptor/Gq Complex. *Nat Commun* **2021**, *12* (1), 2086. <https://doi.org/10.1038/s41467-021-22427-2>.

(59) Bauer, P.; Hess, B.; Lindahl, E. GROMACS 2022 Source Code. Zenodo 2022.

(60) Wieland, K.; Laak, A. M. Ter; Smit, M. J.; Kühne, R.; Timmerman, H.; Leurs, R. Mutational Analysis of the Antagonist-Binding Site of the Histamine H1 Receptor. *Journal of Biological Chemistry* **1999**, *274* (42), 29994–30000. <https://doi.org/10.1074/jbc.274.42.29994>.

(61) Maeda, S.; Qu, Q.; Robertson, M. J.; Skiniotis, G.; Kobilka, B. K. Structures of the M1 and M2 Muscarinic Acetylcholine Receptor/G-Protein Complexes. *Science (1979)* **2019**, *364* (6440), 552–557. <https://doi.org/10.1126/science.aaw5188>.

(62) Venkatakrisnan, A. J.; Fonseca, R.; Ma, A. K.; Hollingsworth, S. A.; Chemparathy, A.; Hilger, D.; Kooistra, A. J.; Ahmari, R.; Babu, M. M.; Kobilka, B. K.; Dror, R. O. Uncovering Patterns of Atomic Interactions in Static and Dynamic Structures of Proteins. November 13, 2019. <https://doi.org/10.1101/840694>.

(63) García-Nafría, J.; Lee, Y.; Bai, X.; Carpenter, B.; Tate, C. G. Cryo-EM Structure of the Adenosine A_{2A} Receptor Coupled to an Engineered Heterotrimeric G Protein. *Elife* **2018**, *7*. <https://doi.org/10.7554/eLife.35946>.

(64) Bian, C.; Xu, C.; Ruan, J.; Lee, K. K.; Burke, T. L.; Tempel, W.; Barsyte, D.; Li, J.; Wu, M.; Zhou, B. O.; Fleharty, B. E.; Paulson, A.; Allali-Hassani, A.; Zhou, J.-Q.; Mer, G.; Grant, P. A.; Workman, J. L.; Zang, J.; Min, J. Sgf29 Binds Histone H3K4me_{2/3} and Is Required for SAGA Complex Recruitment and Histone H3 Acetylation. *EMBO J* **2011**, *30* (14), 2829–2842. <https://doi.org/10.1038/emboj.2011.193>.

(65) Xiong, J.-P.; Mahalingham, B.; Alonso, J. L.; Borrelli, L. A.; Rui, X.; Anand, S.; Hyman, B. T.; Rysiok, T.; Müller-Pompalla, D.; Goodman, S. L.; Arnaout, M. A. Crystal Structure of the Complete Integrin AVβ₃ Ectodomain plus an α/β Transmembrane Fragment. *Journal of Cell Biology* **2009**, *186* (4), 589–600. <https://doi.org/10.1083/jcb.200905085>.

(66) van Niel, M. B.; Fauber, B. P.; Cartwright, M.; Gaines, S.; Killen, J. C.; René, O.; Ward, S. I.; de Leon Boenig, G.; Deng, Y.; Eidenschenk, C.; Everett, C.; Gancia, E.; Ganguli, A.; Gobbi, A.; Hawkins, J.; Johnson, A. R.; Kiefer, J. R.; La, H.; Lockey, P.; Norman, M.; Ouyang, W.; Qin, A.; Wakes, N.; Waszkowycz, B.; Wong, H. A Reversed Sulfonamide Series of Selective ROR_c Inverse Agonists. *Bioorg Med Chem Lett* **2014**, *24* (24), 5769–5776. <https://doi.org/10.1016/j.bmcl.2014.10.037>.

(67) Wagner, J.; von Matt, P.; Sedrani, R.; Albert, R.; Cooke, N.; Ehrhardt, C.; Geiser, M.; Rummel, G.; Stark, W.; Strauss, A.; Cowan-Jacob, S. W.; Beerli, C.; Weckbecker, G.; Evenou, J.-P.; Zenke, G.; Cottens, S. Discovery of 3-(1 H -Indol-3-Yl)-4-[2-(4-

Methylpiperazin-1-Yl)Quinazolin-4-Yl]Pyrrole-2,5-Dione (AEB071), a Potent and Selective Inhibitor of Protein Kinase C Isozymes. *J Med Chem* **2009**, *52* (20), 6193–6196. <https://doi.org/10.1021/jm901108b>.

(68) Maeda, S.; Koehl, A.; Matile, H.; Hu, H.; Hilger, D.; Schertler, G. F. X.; Manglik, A.; Skiniotis, G.; Dawson, R. J. P.; Kobilka, B. K. Development of an Antibody Fragment That Stabilizes GPCR/G-Protein Complexes. *Nat Commun* **2018**, *9* (1), 3712. <https://doi.org/10.1038/s41467-018-06002-w>.

(69) Bacot-Davis, V.; Palmenberg, A. Nucleotide-Free Human Ran Gtpase. October 22, 2014. <https://doi.org/10.2210/pdb2mmc/pdb>.

(70) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM Database and PPM Web Server: Resources for Positioning of Proteins in Membranes. *Nucleic Acids Res* **2012**, *40* (D1), D370–D376. <https://doi.org/10.1093/nar/gkr703>.

(71) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys J* **2009**, *97* (1), 50–58. <https://doi.org/10.1016/j.bpj.2009.04.013>.

(72) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J Comput Chem* **2008**, *29* (11), 1859–1865. <https://doi.org/10.1002/jcc.20945>.

(73) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded

and Intrinsically Disordered Proteins. *Nat Methods* **2017**, *14* (1), 71–73.

<https://doi.org/10.1038/nmeth.4067>.

(74) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Phys* **1981**, *52* (12), 7182–7190.

<https://doi.org/10.1063/1.328693>.

(75) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, *98* (12), 10089–10092.

<https://doi.org/10.1063/1.464397>.

Funding / Acknowledgement

This work was funded by grants from the National Institutes of Health R01-GM117923 to N. V., R01- LM013876 to N. V., A. R., Sergio Branciamore, and Supriyo Bhattacharya, and R01- LM013138 to A. R. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additional support is acknowledged by Dr Susumu Ohno Chair in Theoretical Biology (held by A. R.), and Susumu Ohno Distinguished Investigator Fellowship (to G. G.).