JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

# Measuring Differences in Protein Allosteric Graphs Constructed via Molecular Dynamics Simulations

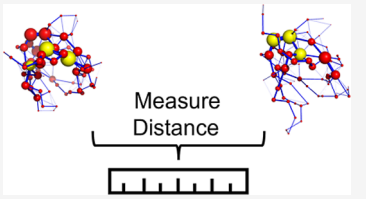
*Published as part of Journal of Chemical Theory and Computation special issue "Modeling the Impact of Protein Mutations on Chemical Reactions for Biochemistry, Protein Engineering, and Drug Discovery".*

Jiale Shi, Zhongyi Wan, Renjie Zhu, and Qiang Cui*

Cite This: https://doi.org/10.1021/acs.jctc.6c00082 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Protein allostery plays essential roles in many biological processes through long-range communication between distant sites. Three-dimensional Shortest Path Map (3D SPM) graphs generated from molecular dynamics trajectories provide an effective, structure-embedded representation of protein conformational dynamics and allosteric communication pathways. However, comparisons among 3D SPM graphs have remained largely qualitative, limiting quantitative insight. Here, we address this gap by introducing an algorithm that uses the earth mover's distance (EMD) and the normalized graph Laplacian to quantitatively measure distances between 3D SPM graphs by evaluating their spatial distributions and network connectivity. This approach allows us to establish the sensitivity of computed allosteric communication networks to the choice of force field and other parameters in the 3D SPM calculations. Furthermore, the quantitative comparison of 3D SPM graphs enables distinguishing mutations that minimally perturb allosteric communication from those that substantially rewire the network. Importantly, decomposition of the EMD highlights the residues and edges that contribute the most to these shifts. This work has the potential to facilitate the analysis of high-throughput experiments regarding protein function and evolution, and to provide guidance for protein engineering.




## 1. INTRODUCTION

Allostery is a regulatory process in biomolecules whereby an event at an allosteric site propagates its effect to a distal site, leading to changes in conformation and/or activity.[1−22] Allosteric mechanisms are ubiquitous; cooperativity can drive conformational changes that span multiple domains and underlie diverse biological processes, including binding cooperativity,[6,23−27] metabolic enzyme regulation,[28−30] signal transduction,[31−36] immune recognition and activation,[13,37−39] and membrane transport.[40−42] Therefore, elucidating and controlling allostery is of major interest for both fundamental mechanistic analysis and practical applications such as protein engineering[24,29,38] and drug design.[14,18,40] However, understanding allosteric regulation is challenging because of the highly decentralized nature and scope of allosteric interactions. Indeed, structural and dynamical perturbations propagate through proteins via conformational rearrangements governed by an intricate interplay between local interactions and collective motions across multiple time scales.[1,3,4,43−47]

Apart from experimental methodologies such as NMR,[8−10] Cryo-EM,[48−50] and Hydrogen−Deuterium Exchange Mass Spectrometry (HDX-MS),[51−53] computational methodologies play a critical role in elucidating allosteric mechanisms. A series of computational and theoretical methods have been developed for allosteric community network analysis, such as the shortest path map (SPM) developed by Osuna et al.,[54−58] MDPath developed by Doering et al.,[102] residue interaction networks (RINs) developed by Yehorova et al.,[103] distal mutations rewire allosteric networks developed by Wang et al.,[104] dynamical network models of allostery developed by Amaro et al.,[11,59,60] dynamical perturbation networks developed by Gheeraert et al.,[105] difference contact network analysis (dCNA) path analysis developed by Hamelberg et al.,[106] dynamical models of allosteric communication based on protein contact clusters developed by Stock et al.,[107] and nonequilibrium molecular dynamics simulations developed by Chan et al.[108] Among those, SPM has proven effective for qualitative analysis across case studies in steering catalysis,[54,55,61] such as retro-aldolase,[58] monoamine oxidase[57,62] and tryptophan synthase.[61,63,64] SPM utilizes Dijkstra's algorithm to generate a residue communication network superimposed on the actual protein 3D structure, forming a 3D SPM graph, which can provide qualitative visual inspection about protein conformational dynamics and allosteric community networks. The 3D SPM graphs facilitate the identification of conformationally influential residues,

Received: January 16, 2026
Revised: April 22, 2026
Accepted: April 28, 2026

including remote positions strongly coupled to the active site. SPM results have overlapped with, and rationalized distal hotspots uncovered by laboratory evolution, reinforcing the role of distributed communication.

However, a principled algorithm for accurately quantifying and measuring differences among 3D SPM graphs is lacking. Establishing such an algorithm is important for evaluating the sensitivity of results to technical details such as force-field, and for enabling quantitative characterization of how mutations reshape allosteric community networks. Here, we address this gap by computing pairwise distances between 3D SPM graphs in both spatial and spectral domains using the earth mover's distance (EMD) and the normalized graph Laplacian, respectively. Moreover, the resulting distance metrics enable identification of residues that undergo major mutation-induced changes in allosteric community networks.

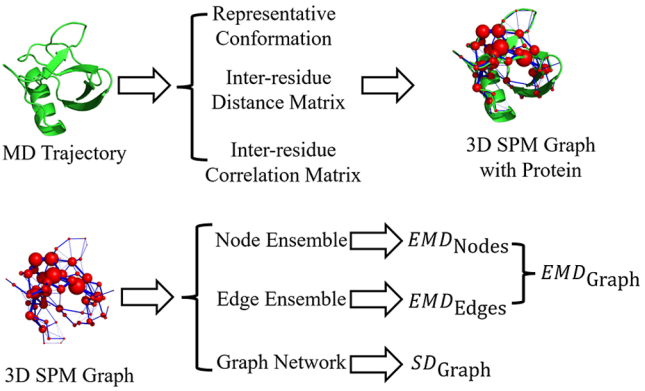



**Figure 1.** To generate a 3D SPM graph, an MD trajectory is processed to select a representative protein conformation (the centroid of the most populated cluster) for 3D placement and to compute the inter-residue distance matrix and displacement-correlation matrix. To quantify the pairwise distance between two 3D SPM graphs, the 3D SPM graphs can be divided into three components: (1) the node ensemble, (2) the edge ensemble, and (3) the graph network without spatial information. The nodes are specified by chemical identity, $C_\alpha$ coordinates, and node significance scores. The edges are specified by midedge coordinates and edge significance scores. Information from the node and edge ensembles is sufficient to reconstruct the 3D SPM graphs. Therefore, combining the node and edge ensemble distance metrics provides a complete measure of difference between two 3D SPM graphs. The distance between two SPM graph networks therefore quantifies differences in the spectral properties of the corresponding 3D SPM graphs.

## 2. METHODS

### 2.1. 3D SPM Graphs

As shown in Figure 1, to generate a 3D SPM graph, an MD trajectory is processed to select a representative protein conformation (the centroid of the most populated cluster) for 3D placement and to compute the inter-residue distance matrix and displacement-correlation matrix, which are the three necessary inputs for SPM calculations. A residue-level graph is then constructed: residues (typically $C_\alpha$) are nodes, and an undirected edge $(i,j)$ is added if $D_{ij} \leq t_{\text{thresh}}$, where $t_{\text{thresh}}$ is the distance threshold. Next, edge lengths are derived from pairwise displacement correlations, via $l_{ij} = -\log|C_{ij}|$. Highly correlated residue

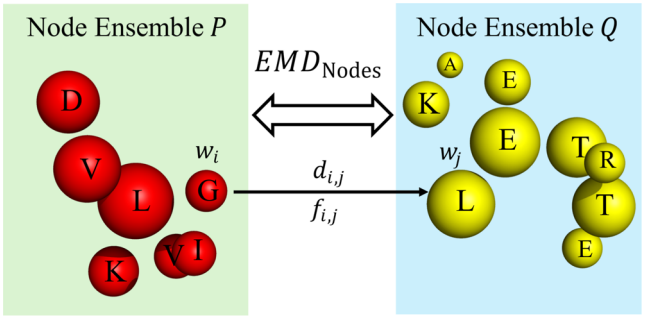



**Figure 2.** Calculation of $EMD_{Nodes}$ between two residue node ensembles.

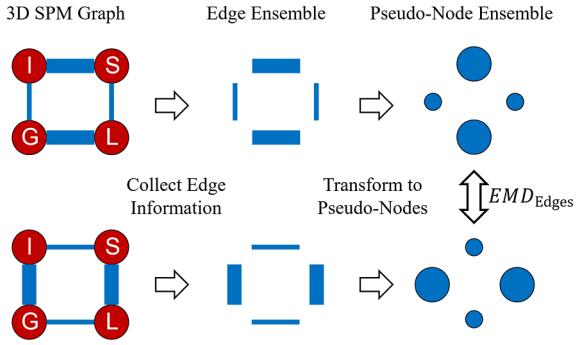



**Figure 3.** Pairwise distance calculation between edge ensembles. First, edge information about the middle point positions and widths are extracted from the 3D SPM graphs. Second, edge ensembles are transformed to be pseudonode ensembles, where the position of the pseudonode is the same as the middle point position of the edge and the score and radius of the pseudonodes is the same as the score and width of the edges. Finally, apply the same EMD calculation algorithm developed in node ensemble distance calculation to calculate $EMD_{Edges}$.

pairs thus receive shorter effective lengths; weakly correlated pairs have longer effective lengths. Using Dijkstra's algorithm,[65] a shortest path is computed between every residue pair in the initial graph. Edge usage frequency across all such paths is tallied and normalized to [0,1] by dividing by the maximum usage across edges, yielding the edge significance score $u_e$. The significance threshold ($s_{\text{thresh}}$) filters graph edges by their score, where only edges with significance score greater than $s_{\text{thresh}}$ are retained. Node significance scores are the sums of scores of incident edges, optionally after applying a significance threshold. Finally, the SPM graphs are mapped onto the representative structure to yield the 3D SPM graphs.[54−58]

To calculate the pairwise distance between two 3D SPM graphs, each graph is decomposed into three components: (1) the node ensemble, (2) the edge ensemble, and (3) the graph network. Nodes are specified by residue chemical identity, $C_\alpha$ coordinates, and node significance scores, which indicate the spatial information on the 3D SPM graphs. Edges are specified by midedge coordinates and edge significance scores, which convery both spatial and spectral information. The information contained in the node and edge ensembles is sufficient to reconstruct the full 3D SPM graphs. Consequently, combining the node- and edge-ensemble distance metrics provides a complete measure of the difference between two 3D SPM graphs. To focus on the topological features of the SPM graphs, the SPM graph network is defined without spatial information about nodes and edges but encodes only spectral (topological) properties. The distance between two SPM graph networks therefore quantifies differences in the spectral properties of the corresponding 3D SPM graphs.

In this work, we adopt the 3D SPM construction following Osuna et al.[54−58] as a concrete and reproducible instantiation. The distance framework itself, however, is not specific to this edge definition and can be applied to any 3D residue graph with well-defined nonnegative node


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

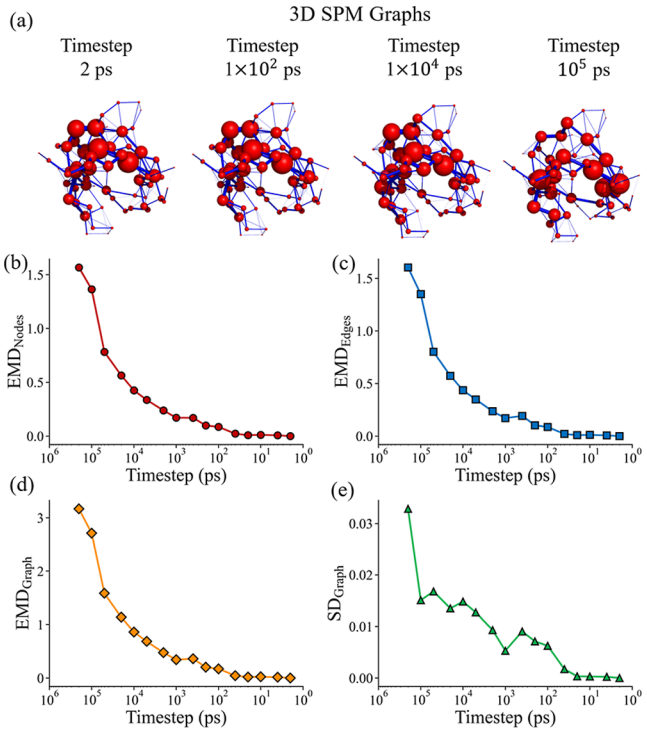



**Figure 4.** (a) 3D SPM graphs generated from a single 1 $\mu$s equilibrium MD trajectory of wild-type ubiquitin using frame intervals of 2 ps, 100 ps, $10^4$ ps, and $10^5$ ps. The 3D SPM graph generated with a 2 ps frame interval is used as the reference to evaluate how the frame interval affects pairwise distances. (b) The pairwise $EMD_{Nodes}$ as a function of frame interval time step. (c) The pairwise $EMD_{Edges}$ as a function of frame interval time step. (d) The pairwise $EMD_{Graph}$ as a function of frame interval time step. (e) The pairwise $SD_{Graph}$ as a function of frame interval time step.

and edge weights. Accordingly, alternative edge-weighting schemes, such as residue−residue contact scores, mutual-information-based couplings, or other communication metrics, can be used as inputs; such substitutions would change the resulting distances by changing the underlying graph representation and can therefore be assessed within the same quantitative framework.

## 2.2. Node Ensemble Pairwise Distance

With the node ensembles and the pairwise distance between two single nodes defined (see below), the earth mover's distance (EMD) is a metric that is well-constructed to calculate the similarity of ensembles; it has been successfully applied in multiple fields for ensemble similarity calculation, such as the similarity of inorganic solids,[66,67] the similarity between complex polymers,[68,69] the similarity of biomarker expression levels,[109] and geometric data set distances.[110] EMD may be conceptualized as the minimal amount of work to transform one distribution into another, and it can be formulated and solved as a transportation problem. Here, the problem is transforming one discrete weighted node ensemble $P$ to another $Q$ with the minimum amount of work done, which can be interpreted as a measure of distance.

Thus, the next step is to determine the pairwise distance ($d_{Node}$) of the individual nodes of the SPM graphs. The residue nodes in SPM graphs have both spatial locations and chemical identities. Therefore, $d_{Node}$ includes two aspects: the spatial distance ($d_{3D}$) and the chemical distance ($d_{Chem}$).

$$d_{Node} = d_{3D} + \alpha \cdot d_{Chem} \tag{1}$$

here $\alpha$ is a parameter to control the relative weight ratio between the spatial distance and the chemical distance, with the default value of 1. The chemical-distance term is included because protein structure is often more conserved than protein sequence, multiple mutations may occur without substantially altering the overall fold and the resulting spatial relationships. In such cases, chemical distance can systematically improve the sensitivity of SPM pairwise distance calculations, while a purely spatial metric may not adequately discriminate among variants. Users can set $\alpha = 0$ for purely spatial comparisons or increase $\alpha$ when greater emphasis on physicochemical differences is desired. $d_{3D}$ is calculated by taking the spatial locations of the two residue nodes.$d_{Chem}$ is calculated by the Tanimoto distance by using chemical informatics based on the chemical identities of the two residue nodes. First, each node residue is represented by a SMILES (Simplified Molecular-Input Line-Entry System) string.[70,71] For example, the serine (Ser) residue is "*N[C@@H](CO)C(−O)*". The symbol (*) is used as a connection point to clearly illustrate which atoms of the residue are connected in the protein main chains. Modified residues can also be represented; for instance, phosphoserine (pSer) is "*N[C@@H](COP(=O)(O)O)-C(−O)*", and the doubly deprotonated form ($pSer^{2-}$) is "*N[C@@H](COP(=O)([O-])[O-])C(−O)*". SMILES representations capture not only the residue's chemical structure but also its protonation state. Morgan fingerprints with Tanimoto distance as a conservative default for $d_{Chem}$ calculation because this combination is widely used and well characterized for molecular similarity comparisons, with strong support in common cheminformatics toolkits. These SMILES strings are then transformed into fingerprints using Morgan fingerprints bit vector (radius = 2, nBits = 2048) as implemented in RDKit.[111] Then, the pairwise chemical distance ($d_{Chem}$) between the node residues $p_i$ and $q_j$ is calculated using the Tanimoto distance metric.[112] The chemical distance score $d_{Chem}$ ranges from 0 to 1, where self-distance is 0. Greater chemical dissimilarity between two residues yields a larger chemical distance $d_{Chem}$.

In SPM 3D graphs, node size encodes significance scores ($r_i \propto score_i$). We assign each residue node $i$ a weight $w_i$ proportional to $score_i$, $w_i \propto r_i \propto score_i$, thus larger (more significant) nodes receive higher weights in the EMD calculation. Weights of nodes in one graph are then normalized so that their sum over all nodes in a given protein 3D SPM graph equals 1.

After all necessary information is obtained on $w_{pi}$, $w_{qj}$ and $d_{i,j}$ for all the nodes in the ensembles, the optimized transport flows $F = [f_{i,j}]$ and the EMD are determined using eq 2a along with the constraints specified in eqs 2b−e.

$$\mathrm{EMD}(P, Q) = \frac{\min_F \sum_{i=1}^{m}\sum_{j=1}^{n}(f_{i,j}\cdot d_{i,j})}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j}} = \min_F \sum_{i=1}^{m}\sum_{j=1}^{n}(f_{i,j}\cdot d_{i,j}) \tag{2a}$$

$$\text{Subject to } f_{i,j} \geq 0, \quad \text{for any } 1 \leq i \leq m, \quad 1 \leq j \leq n \tag{2b}$$

$$\sum_{j=1}^{n} f_{i,j} = w_{p_i}, \quad \text{for any } 1 \leq i \leq m \tag{2c}$$

$$\sum_{i=1}^{m} f_{i,j} = w_{q_j}, \quad \text{for any } 1 \leq j \leq n \tag{2d}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{i,j} = \sum_{i=1}^{m} w_{p_i} = \sum_{j=1}^{n} w_{q_j} = 1 \tag{2e}$$

$f_{i,j}$ represents the amount of weight at $p_i$ that is transported to $q_j$. The sum of all the individual flows originating from $p_i$ is equal to the weight $w_{pi}$, and equivalently, the sum of all the individual flows originating from $q_j$ is equal to the weight $w_{qj}$, shown in eqs 2c,d. Here, $f_{i,j}\cdot d_{i,j}$ is the cost for each individual flow. Thus, EMD represents the minimum overall cost to convert one node ensemble $P$ to another node ensemble $Q$.

To localize the drivers of node ensemble distances, we decompose the EMD into per-node contributions. Given the optimal transport plan $F = [f_{ij}]$ and ground distances $d_{ij}$ between node $i$ in system $P$ and $j$ in system $Q$, the contribution of node $i$ in $P$ is $\sum_{j=1}^{n} f_{i,j}\cdot d_{i,j}$, and the contribution of node $j$ in $Q$ is $\sum_{i=1}^{m} f_{i,j}\cdot d_{i,j}$. Ranking $\sum_{j=1}^{n} f_{i,j}\cdot d_{i,j}$ and


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

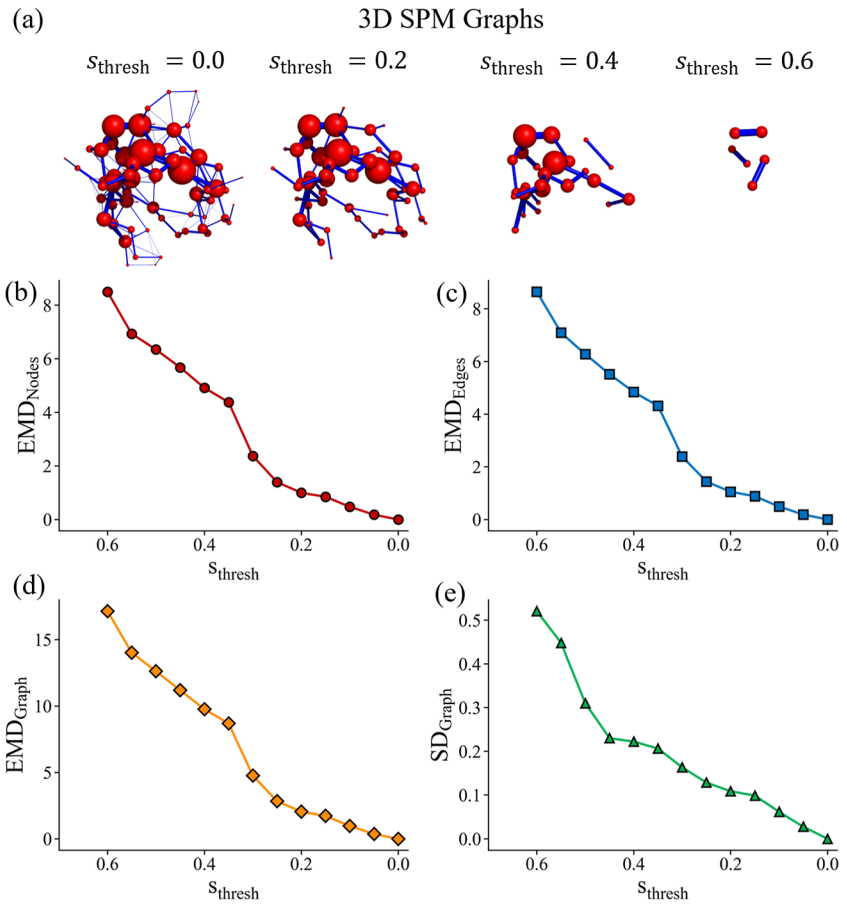



**Figure 5.** (a) 3D SPM graphs generated from a single 1 $\mu$s equilibrium MD trajectory of wild-type ubiquitin using $s_{thresh}$ of 0.0, 0.2, 0.4, and 0.6. The 3D SPM graph generated with $s_{thresh}$ = 0.0 is used as the reference to evaluate how $s_{thresh}$ affects pairwise distances. (b) The pairwise $EMD_{Nodes}$ as a function of $s_{thresh}$. (c) The pairwise $EMD_{Edges}$ as a function of $s_{thresh}$. (d) The pairwise $EMD_{Graph}$ as a function of $s_{thresh}$. (e) The pairwise $SD_{Graph}$ as a function of $s_{thresh}$.

**Table 1. Number of Nodes and Number of Edges for Different $s_{thresh}$**[a]

| | number of nodes | | | | number of edges | | | |
|---|---|---|---|---|---|---|---|---|
| $s_{thresh}$ | score > 0.0 | score > 0.1 | score > 0.2 | score > 0.3 | score > 0.0 | score > 0.1 | score > 0.2 | score > 0.3 |
| 0.0 | 76 | 76 | 65 | 60 | 163 | 90 | 60 | 40 |
| 0.05 | 75 | 74 | 64 | 59 | 112 | 90 | 60 | 40 |
| 0.1 | 69 | 69 | 59 | 56 | 90 | 90 | 60 | 40 |
| 0.15 | 59 | 59 | 56 | 49 | 70 | 70 | 60 | 40 |
| 0.2 | 56 | 56 | 56 | 47 | 60 | 60 | 60 | 40 |
| 0.25 | 51 | 51 | 51 | 46 | 53 | 53 | 53 | 40 |
| 0.3 | 42 | 42 | 42 | 42 | 40 | 40 | 40 | 40 |
| 0.35 | 28 | 28 | 28 | 28 | 24 | 24 | 24 | 24 |
| 0.4 | 25 | 25 | 25 | 25 | 21 | 21 | 21 | 21 |
| 0.45 | 19 | 19 | 19 | 19 | 16 | 16 | 16 | 16 |
| 0.5 | 15 | 15 | 15 | 15 | 11 | 11 | 11 | 11 |
| 0.55 | 13 | 13 | 13 | 13 | 7 | 7 | 7 | 7 |
| 0.6 | 6 | 6 | 6 | 6 | 3 | 3 | 3 | 3 |

[a]For the node score distribution and edge score distribution with different $s_{thresh}$, see Figures S1−S13 in the Supporting Information.

$\sum_{i=1}^{m} f_{i,j} \cdot d_{i,j}$ identifies the residues that contribute most to the observed node ensemble distances.

These equations are coded into POT,[72,73] an open source Python library that provides solvers for optimization problems related to optimal transport for signal, image processing and machine learning, to solve the EMD between two node ensembles.

Apart from the default setting, other fingerprint representations, including MACCS and RDKit topological fingerprints, as well as other similarity functions, including Dice and cosine are also available for users to select. An analysis of the effects of fingerprint representations and similarity functions on $EMD_{Node}$ for single-mutation pairwise distance analysis is included in the Supporting Information (Table S1). Enabling the chemical distance term (set $\alpha$ = 1) increased $EMD_{Node}$ by 2.41 to 5.08% relative to the spatial-only setting (set $\alpha$ = 0) for all fingerprint/similarity combinations tested. Among the fingerprint/similarity combinations tested, Morgan fingerprint and Tanimoto distance generally produce the largest enhancement relative to the


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

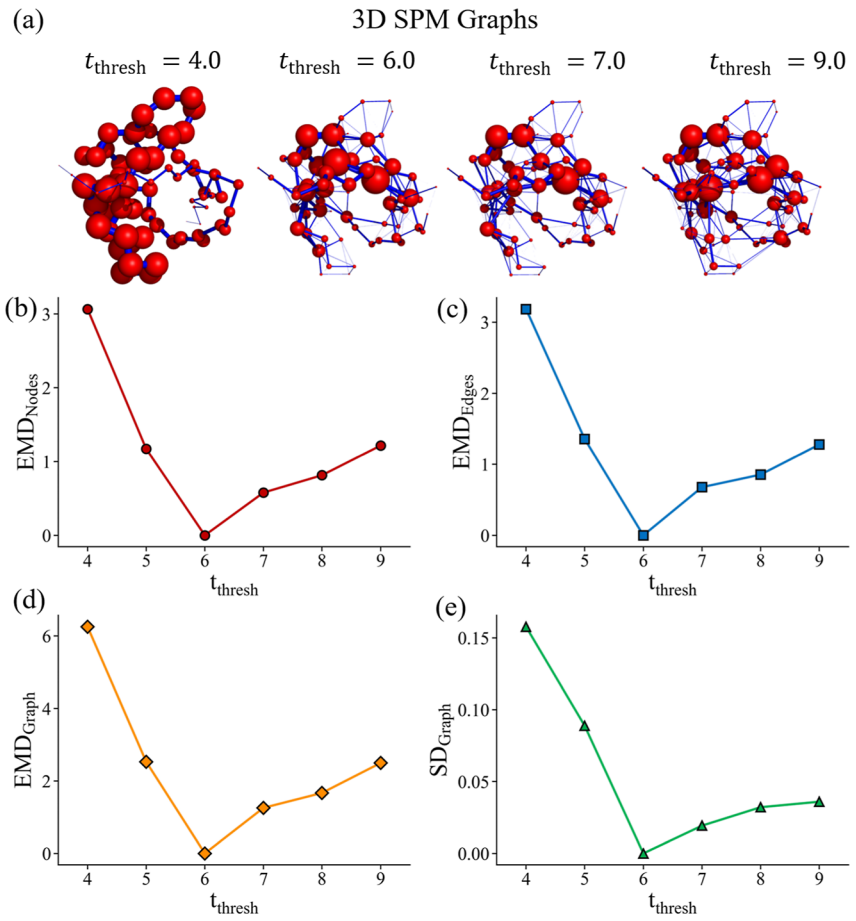



**Figure 6.** (a) 3D SPM graphs generated from a single 1 $\mu$s equilibrium MD trajectory of wild-type ubiquitin using $t_{thresh}$ of 4, 6, 7, and 9. The 3D SPM graph generated with $t_{thresh}$ = 6 is used as the reference to evaluate how $t_{thresh}$ affects pairwise distances. (b) The pairwise $EMD_{Nodes}$ as a function of $t_{thresh}$. (c) The pairwise $EMD_{Edges}$ as a function of $t_{thresh}$. (d) The pairwise $EMD_{Graph}$ as a function of $t_{thresh}$. (e) The pairwise $SD_{Graph}$ as a function of $t_{thresh}$.

**Table 2. Number of Nodes and Number of Edges for Different $t_{thresh}$**[a]

| | number of nodes | | | | number of edges | | | |
|---|---|---|---|---|---|---|---|---|
| $t_{thresh}$ | score > 0.0 | score > 0.1 | score > 0.2 | score > 0.3 | score > 0.0 | score > 0.1 | score > 0.2 | score > 0.3 |
| 4 | 76 | 74 | 72 | 70 | 75 | 73 | 67 | 63 |
| 5 | 76 | 72 | 61 | 56 | 95 | 73 | 54 | 42 |
| 6 | 76 | 76 | 65 | 60 | 163 | 90 | 60 | 40 |
| 7 | 76 | 76 | 66 | 61 | 201 | 86 | 53 | 25 |
| 8 | 76 | 76 | 67 | 62 | 212 | 92 | 50 | 36 |
| 9 | 76 | 76 | 67 | 62 | 226 | 97 | 55 | 34 |

[a]For the node score distribution and edge score distribution with different $t_{thresh}$, see Figures S14–S19 in the Supporting Information.

spatial-only setting, while the overall variant ranking remained stable across representations. These results indicate that under the single mutation, the 3D term is the primary determinant of the overall transport distance, whereas the chemical term provides a modest but consistent enhancement in sensitivity, refining it by increasing separation among related protein variants under a chemically informed weighting scheme. Including chemical information is therefore not intended to override geometric matching, but to provide an additional constraint when spatial information alone is insufficient to distinguish plausible correspondences.

For same-sequence comparisons, residue-level interpretation is supported by an explicit structural alignment step performed before graph construction and EMD calculation. When the overall fold is conserved, this alignment makes corresponding residues spatially proximal, such that the optimal transport plan is effectively biased toward local matching. By contrast, for proteins with substantial sequence or structural differences, the EMD formulation is not intended to enforce residue-index-preserving correspondence, but rather to provide a flexible distributional comparison between graphs.

## 2.3. Edge Ensemble Pairwise Distance

A node-ensemble distance alone cannot fully capture differences between 3D SPM graphs: in principle (see Figures 2 and 3), two SPMs can share the same node ensemble yet differ in their connectivity. To quantify such differences, we compute a pairwise distance between edge ensembles using the same EMD framework applied to nodes.

First, edge information about middle point positions and widths are extracted from the 3D SPM graphs. Subsequently, edge ensembles are transformed into pseudonode ensembles, where the position of the pseudonode is given by the middle point position of the edge; the radius of the pseudonode is given by the width of the edge. Finally, apply the same EMD calculation algorithm used in node ensemble distance calculation to calculate $EMD_{Edges}$.


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

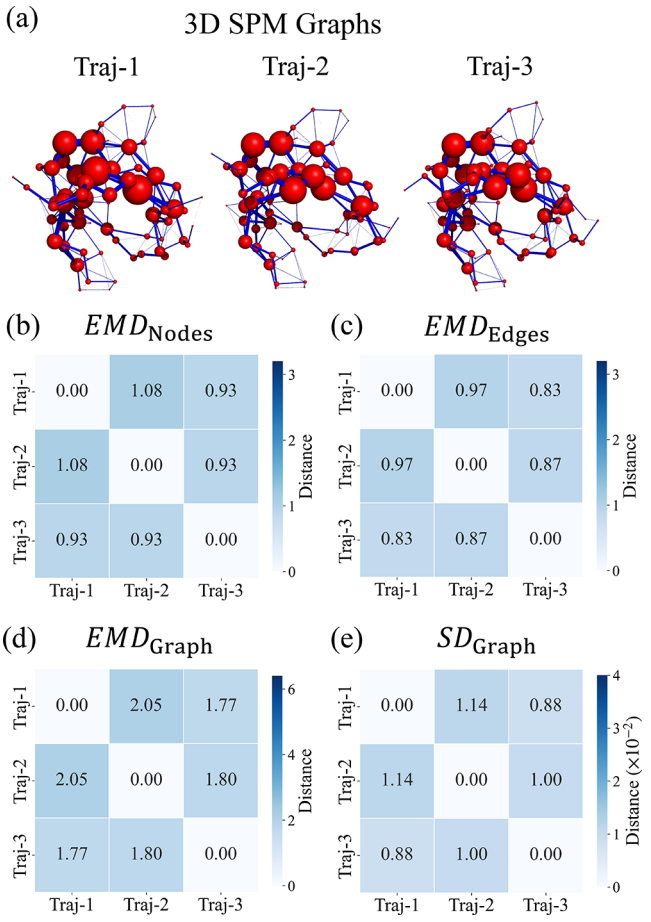



**Figure 7.** (a) Three 3D SPM graphs are generated from three independent 1 $\mu$s equilibrium MD trajectories (Traj-1, Traj-2, and Traj-3) of wild-type ubiquitin using the same CHARMM36m force field with different initial configurations. The large nodes and thick edges remain largely consistent, whereas the small nodes and thin edges exhibit variability. (b) The pairwise $EMD_{Nodes}$ matrix for Traj-1, Traj-2 and Traj-3. (c) The pairwise $EMD_{Edges}$ matrix. (d) The pairwise $EMD_{Graph}$ matrix. (e) The pairwise $SD_{Graph}$ matrix.

## 2.4. Overall Pairwise EMD Between 3D SPM Graphs

Two EMD scores are obtained: $EMD_{Nodes}$ for node ensembles and $EMD_{Edges}$ for edge ensembles. The overall pairwise $EMD_{Graph}$ between two 3D SPM graphs is taken as a linear combination of the $EMD_{Nodes}$ and $EMD_{Edges}$

$$EMD_{graph} = \beta_1 \cdot EMD_{Nodes} + \beta_2 \cdot EMD_{Edges} \quad (3)$$

here, $\beta_1$ and $\beta_2$ are tunable parameters that control the relative contributions of the node and edge terms, respectively, with default values of 1. Equal default weights were chosen as a simple and neutral baseline in the absence of an application-specific prior favoring either node-level or edge-level differences. Under the default setting, $EMD_{Nodes}$ includes both node-level spatial and chemical contributions, whereas $EMD_{Edges}$ reflects edge-level spatial differences. Thus, $EMD_{Graph}$ captures complementary node- and edge-based differences between 3D SPM graphs. Users may adjust $\beta_1$ and $\beta_2$ when they wish to place greater emphasis on node-driven or edge-driven discrepancies. A sensitivity analysis of $\beta_1$ and $\beta_2$ is provided in the Supporting Information.

## 2.5. Spectral Distance

The spectral distance ($SD_{Graph}$) between two weighted graphs provides a permutation-invariant measure of structural similarity based on differences in the eigenvalue spectra of their normalized Laplacian matrices. Because eigenvalues are unchanged under relabeling of the nodes, this metric captures intrinsic topological features of the graph. In the present context, the spectral distance isolates differences in graph connectivity without incorporating 3D spatial coordinates, thereby enabling quantitative comparison of 3D SPM-derived graphs solely on the basis of their network structure.

The normalized Laplacian[74−76] ($L_{norm}$) of a graph is defined as

$$L_{norm} = I - D^{-1/2} W D^{-1/2} \quad (4)$$

$I$ represents the identity matrix. $W$ denotes the weighted adjacency matrix, where $W_{ij}$ represents the weight of edge between nodes $i$ and $j$. $D$ is the diagonal degree matrix of a graph, where each diagonal entry $D_{ij}$ is the sum of the weights of all edges connected to node $i$. $D^{1/2}$ is the inverse square root of $D$, used for normalization. The eigenvalues of $L_{norm}$ (one per node) are computed for each graph, and each eigenvalue is assigned a uniform weight of $\frac{1}{N}$, where $N$ is the number of nodes in the respective graph. This normalization ensures that graphs of different sizes yield spectral distributions with equal total mass. The resulting sets of eigenvalues and associated weights define one-dimensional discrete spectral distributions for each graph.

The distance between two such spectral distributions is then quantified using the first Wasserstein distance, computed with the Wasserstein distance function implemented in SciPy. The resulting scalar value, referred to here as $SD_{Graph}$, reflects the degree of distance between the two spectral distributions and thereby provides a quantitative measure of the difference in spectral features between two 3D SPM graphs.

The distance formulation based on EMD and the normalized graph Laplacian does not require the compared 3D SPM graphs to have the same number of nodes or edges and can therefore accommodate proteins with insertions, deletions, or different graph sizes. When the overall fold is conserved, the resulting distance can be interpreted more directly in terms of localized differences, whereas for larger structural differences it is more appropriately viewed as a global graph-level dissimilarity.

## 2.6. Data and Simulation

To illustrate the distance-quantification algorithm for 3D SPM graphs, we carried out and assembled a series of molecular dynamics (MD) simulations.

(1) Independent replicate simulations of ubiquitin[77] (PDB: 1UBQ): Three independent MD simulations of wide-type ubiquitin were initiated from distinct starting conformations using CHARMM36m,[78] yielding three replicate independent trajectories to estimate the replicate-level variation.
(2) Force-field comparison of ubiquitin[77] (PDB: 1UBQ): Wild-type ubiquitin trajectories were generated with three force fields including CHARMM36m,[78] CHARMM36[79] and AMBER-ff19SB,[80] to assess force-field effects on the distance metrics.
(3) Mutation comparison of ubiquitin[77] (PDB: 1UBQ): Trajectories were obtained for WT and six variants (I3A, L15S, K33G, P38Q, E51D and V70N), using CHARMM36m[78] to quantify mutation effects on 3D SPM graphs.
(4) Mutation comparison of bacterial tetracycline repressor[81,82] (TetR): Trajectories were obtained for WT and two variants (G102D and E150K/C195Q), using CHARMM36m[78] to quantify mutation effects on 3D SPM graphs.

All-atom MD simulations were generated using the Chemistry at Harvard Macromolecular Mechanics (CHARMM)[83,84] graphical user interface (CHARMM-GUI)[85−89] and GROMACS molecular dynamics simulation package. The details of the MD simulations are included in the Supporting Information. For ubiquitin (WT and its six variants), a 2 $\mu$s equilibration run was performed in the isothermal−isobaric (NPT) ensemble using a 2 fs time step at 303.15 K, followed by a 1 $\mu$s production simulation. The 1 $\mu$s production trajectories were used to compute the representative protein conformation, the average inter-residue distance and correlation matrices, and subsequently to construct the 3D SPM graphs of ubiquitin. For TetR (WT and its two variants), which constitute a larger and more complex system than ubiquitin, a 3 $\mu$s equilibration run was performed under the same NPT


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

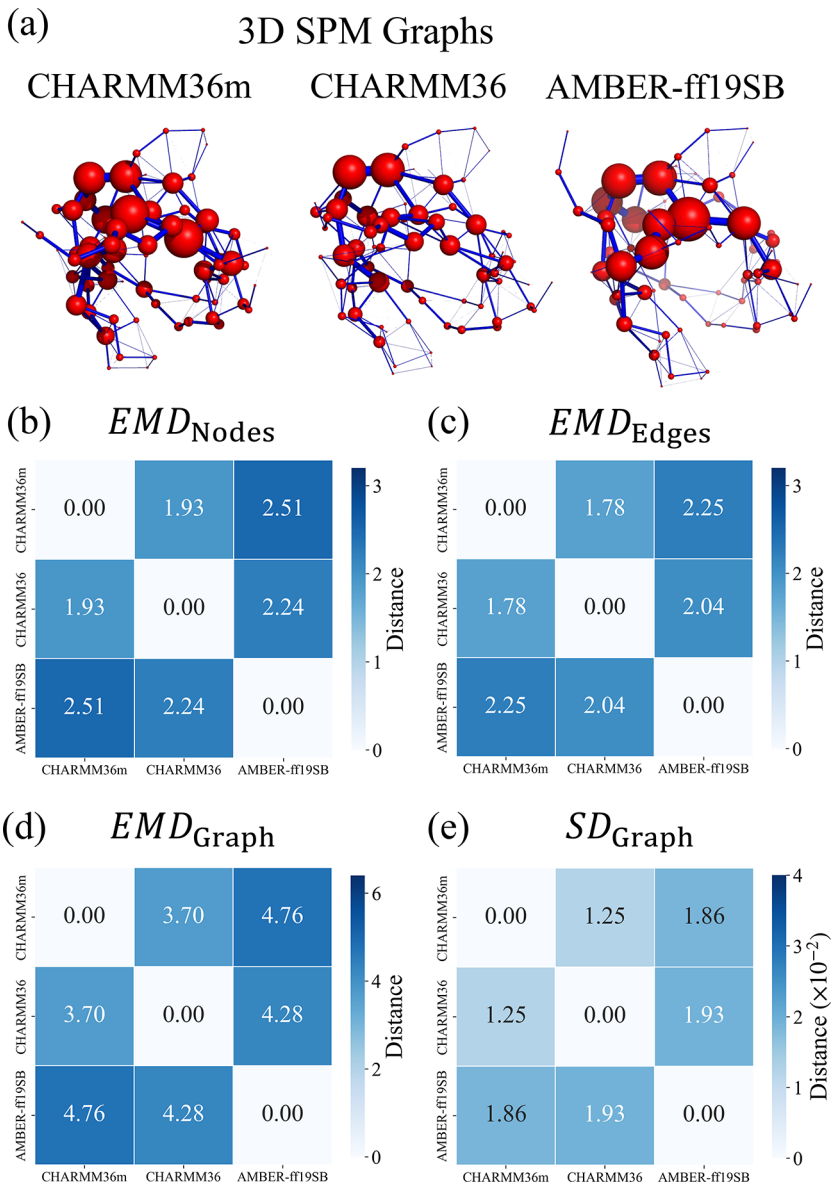



**Figure 8.** (a) Three 3D SPM graphs are generated from equilibrium MD trajectories simulated with three different force fields (CHARMM36m, CHARMM36, and AMBER-ff19SB). (b) The pairwise $EMD_{Nodes}$ matrix for CHARMM36m, CHARMM36, and AMBER-ff19SB. (c) The pairwise $EMD_{Edges}$ matrix. (d) The pairwise $EMD_{Graph}$ matrix. (e) The pairwise $SD_{Graph}$ matrix.

conditions, followed by a 2 $\mu$s production simulation. The 2 $\mu$s production trajectory was used to compute the representative protein conformation, the average inter-residue distance and correlation matrices, and to construct the 3D SPM graphs of TetR. Because the two TetR chains are identical, the corresponding average inter-residue distance and correlation matrices should in principle be the same. In practice, small differences arose from simulation fluctuations. Therefore, a symmetrization procedure was applied to enforce identical average inter-residue distance and correlation matrices for the two chains.

When multiple simulations represent independent replicates of the same physical ensemble of proteins, their trajectories may be combined to improve sampling for construction of the distance and correlation matrices. By contrast, simulations sampling distinct conformational states or performed under different physical conditions should generally be analyzed separately to avoid averaging over state-specific communication patterns.

## 3. RESULTS AND DISCUSSIONS

In the following, we first discuss the sensitivity of computed 3D SPM graph to control parameters such as the number of trajectory frames and the distance threshold for edge detection, then we explore the variation of computed 3D SPM graph when independent trajectories or different force fields are used. Finally, we dissect how mutations perturb the motional coupling network and illustrate how our approach enables the identification of key residues that undergo large changes in coupling network upon mutation.

### 3.1. Different Trajectory Frame Interval Time Step

In Figure 4, using a single densely saved 1 $\mu$s equilibrium MD trajectory about the wide-type ubiquitin, the trajectory frame interval is varied to assess how the number of MD frames quantitatively influences the resulting 3D SPM graphs; the graph generated at the 2 ps time step is used as the reference. As shown in Figure 4a, the 3D SPM graph generated with a 100 ps time


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

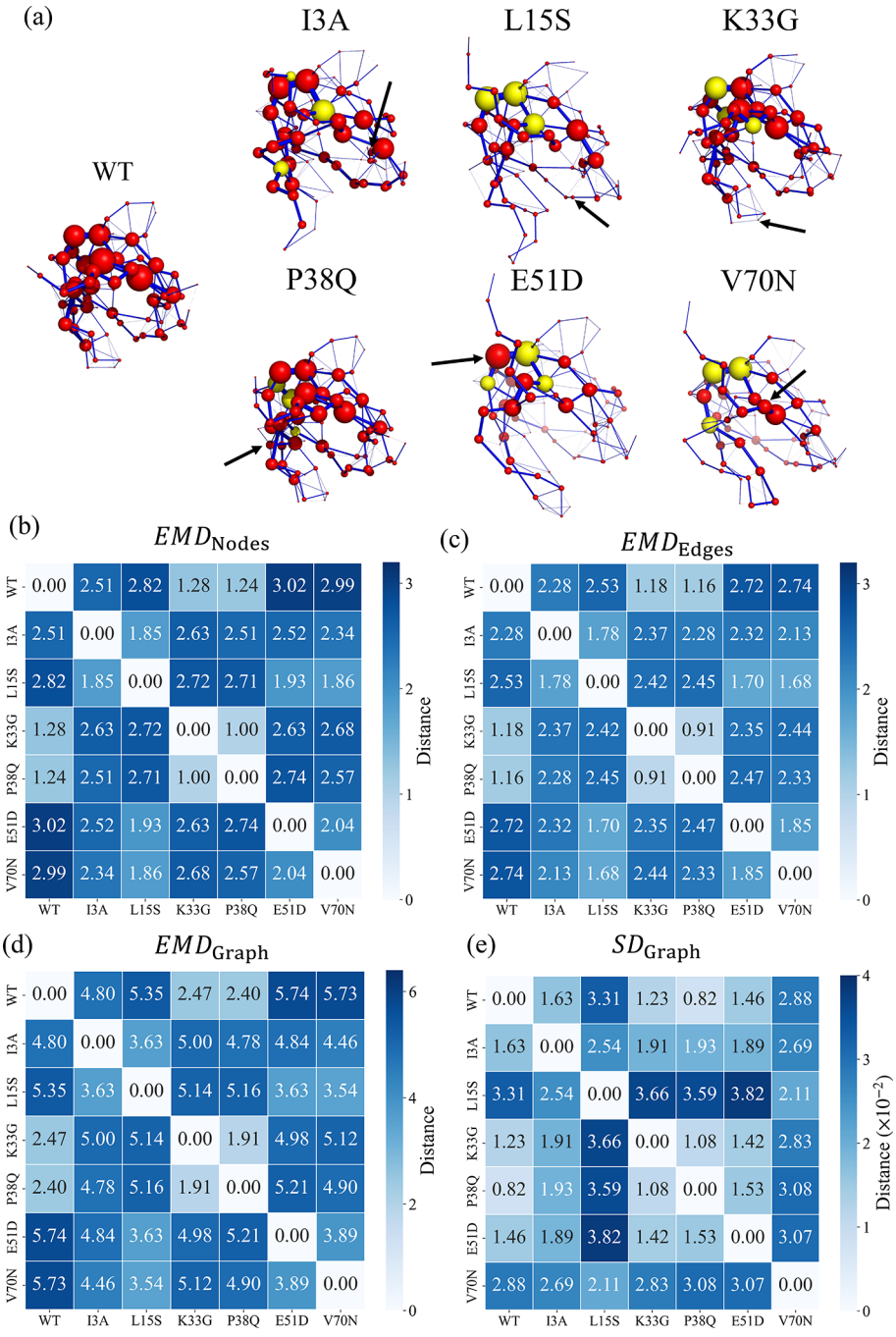



**Figure 9.** (a) Seven 3D SPM graphs are generated from equilibrium MD trajectories for WT ubiquitin and its six variants (I3A, L15S, K33G, P38Q, E51D, and V70N). The location of the mutation is indicated by the arrow. The top three contribution residues for $EMD_{Nodes}$ with WT are marked in yellow (also see Table 3). (b) The pairwise $EMD_{Nodes}$ matrix for WT ubiquitin and its six variants. (c) The pairwise $EMD_{Edges}$ matrix. (d) The pairwise $EMD_{Graph}$ matrix. (e) The pairwise $SD_{Graph}$ matrix.

step is nearly identical to that obtained with a 2 ps time step. When the time step is increased to $10^4$ ps, small deviations begin to appear in small nodes, and at $10^5$ ps the resulting 3D SPM graph shows evident changes. The quantitative changes of the pairwise distances ($EMD_{Nodes}$, $EMD_{Edges}$, $EMD_{Graph}$, $SD_{Graph}$) with time step are shown in Figure 4b–e. As the frame interval decreases from $2 \times 10^5$ ps to $10^4$ ps, the four pairwise distances all drop sharply. When the interval time step is further reduced below $10^4$ ps, the decrease in pairwise distances becomes much more gradual. In particular, for frame intervals smaller than 100 ps, the pairwise distances are nearly zero and remains almost constant. Therefore, based on these observations, a frame interval of $10^4$ ps yields well-converged pairwise distance values, and a frame interval of 100 ps yields strictly converged pairwise distances.

### 3.2. Different Significance Threshold ($s_{thresh}$)

The significance threshold ($s_{thresh}$) is used to filter graph edges based on their scores to balance comprehensive coverage and visual clarity: only edges with score greater than $s_{thresh}$ are retained. Nodes that remain connected through these retained edges are then kept, and their scores are recalculated using only the retained edges. In Figure 5a and Table 1, as $s_{thresh}$ increases,


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

**Table 3. Top Three Contribution Residue Nodes in $EMD_{Nodes}$ Calculation for Ubiquitin Mutation**

| | | first | second | third |
|---|---|---|---|---|
| WT-I3A | WT | LEU73 | ARG42 | LEU69 |
| | | 0.1489 | 0.1270 | 0.1194 |
| | I3A | PRO38 | LEU73 | LEU43 |
| | | 0.1470 | 0.1291 | 0.1053 |
| WT-L15S | WT | GLN40 | ARG42 | LEU69 |
| | | 0.1402 | 0.1257 | 0.1230 |
| | L15S | LEU50 | GLU51 | LEU43 |
| | | 0.2520 | 0.2251 | 0.1356 |
| WT-K33G | WT | LYS6 | VAL5 | LEU69 |
| | | 0.1150 | 0.1033 | 0.0705 |
| | K33G | GLU51 | THR55 | LEU71 |
| | | 0.1112 | 0.0856 | 0.0747 |
| WT-P38Q | WT | HSD68 | THR66 | LEU69 |
| | | 0.1302 | 0.0844 | 0.0807 |
| | P38Q | THR55 | ARG54 | ASP21 |
| | | 0.1188 | 0.1020 | 0.0622 |
| WT-E51D | WT | LEU73 | LEU69 | VAL5 |
| | | 0.1236 | 0.1176 | 0.1157 |
| | E51D | LEU43 | ASP52 | LEU50 |
| | | 0.2550 | 0.2090 | 0.1760 |
| WT-V70N | WT | ARG42 | GLN40 | GLN41 |
| | | 0.2291 | 0.1859 | 0.1404 |
| | V70N | GLU51 | LEU50 | GLU24 |
| | | 0.2305 | 0.2111 | 0.1602 |

more edges and nodes are removed and the graph becomes sparser. For example, when $s_{thresh} = 0.3$, about 25% of the edges are retained. As shown in Figure 5b–d, $EMD_{Nodes}$, $EMD_{Edges}$, and $EMD_{Graph}$, exhibit a pronounced sensitivity to $s_{thresh}$ when $s_{thresh} > 0.3$, whereas for $s_{thresh} < 0.3$, their dependence on $s_{thresh}$ becomes comparatively gradual. $SD_{Graph}$ follows a similar trend, although the transition point occurs at a slightly larger value of $s_{thresh}$. As shown in Figure 5e, when $s_{thresh} > 0.45$, $SD_{Graph}$ changes sharply as a function of $s_{thresh}$, while when $s_{thresh} < 0.45$, the change of $SD_{Graph}$ becomes moderate. Therefore, for qualitative visual comparison, $s_{thresh} = 0.3$ is a reasonable option to balance comprehensive coverage and visual clarity of 3D SPM graphs, consistent with the empirically chosen default value reported by Osuna et al.[54,55] Increasing $s_{thresh}$ further can lead to the serious loss of important information, causing the filtered 3D SPM graphs to become markedly different from the raw graphs without postprocessing. For quantitative comparison, however, to avoid threshold-induced artifacts, all raw edges of the 3D SPM graphs should be retained, without any postprocessing filtering. On the other hand, the algorithm for calculating pairwise distances proposed in this work is computationally efficient to process 3D SPM graphs with all raw edges.

### 3.3. Different Distance Threshold ($t_{thresh}$)

The distance threshold ($t_{thresh}$) is used to restrict which residue pairs are eligible to form edges during the initial graph construction. An undirected edge between residues $i$ and $j$ is added only if the average inter-residue distance is smaller than $t_{thresh}$. As shown in Figure 6a, the choice of $t_{thresh}$ strongly shapes the resulting 3D SPM graphs. Decreasing $t_{thresh}$ prunes contacts and reduces path availability. Consequently, many shortest paths are forced through the same limited set of edges, and the normalized edge-usage scores become narrowly distributed and skewed toward high values (near 1). Nodes incident to these surviving edges appear larger, which can obscure distinctions among the most important allosteric communication routes. Increasing the distance threshold admits more contacts, yielding a denser initial graph with additional alternative routes. This expanded path availability can alter the computed shortest paths and produces more complex 3D SPM graphs. Furthermore, the different slopes of the distance curves for $t_{thresh} > 6$ and $t_{thresh} < 6$

**Table 4. Top Three Contribution Edges in $EMD_{Edges}$ Calculation for Ubiquitin Mutation**

| | | first | second | third |
|---|---|---|---|---|
| WT-I3A | WT | LEU73-ARG74 | ARG72-LEU73 | ARG74-GLY75 |
| | | 0.1576 | 0.1242 | 0.0776 |
| | I3A | GLY35-ILE36 | ILE36-PRO37 | LEU73-ARG74 |
| | | 0.0998 | 0.0957 | 0.0955 |
| WT-L15S | WT | LEU73-ARG74 | ARG72-LEU73 | ARG74-GLY75 |
| | | 0.1227 | 0.0968 | 0.0853 |
| | L15S | LEU50-GLU51 | LEU73-ARG74 | ARG72-LEU73 |
| | | 0.1907 | 0.0992 | 0.0880 |
| WT-K33G | WT | VAL5-LYS6 | VAL5-ILE13 | ILE3-PHE4 |
| | | 0.0678 | 0.0631 | 0.0455 |
| | K33G | LEU50-GLU51 | ARG54-THR55 | THR22-ILE23 |
| | | 0.0837 | 0.0543 | 0.0529 |
| WT-P38Q | WT | LYS6-LEU69 | THR66-LEU67 | GLN41-ARG42 |
| | | 0.0642 | 0.0633 | 0.0617 |
| | P38Q | ARG54-THR55 | THR55-LEU56 | GLU18-ASP21 |
| | | 0.0836 | 0.0624 | 0.0485 |
| WT-E51D | WT | LEU73-ARG74 | ARG72-LEU73 | ARG74-GLY75 |
| | | 0.1142 | 0.0952 | 0.0932 |
| | E51D | LEU50-ASP51 | LEU43-LEU50 | LEU73-ARG74 |
| | | 0.1736 | 0.1295 | 0.1094 |
| WT-V70N | WT | GLN41-ARG42 | LEU73-ARG74 | ARG72-LEU73 |
| | | 0.1591 | 0.1002 | 0.0797 |
| | V70N | GLU24-ASP52 | GLU51-ASP52 | LEU50-GLU51 |
| | | 0.1502 | 0.1424 | 0.1205 |


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

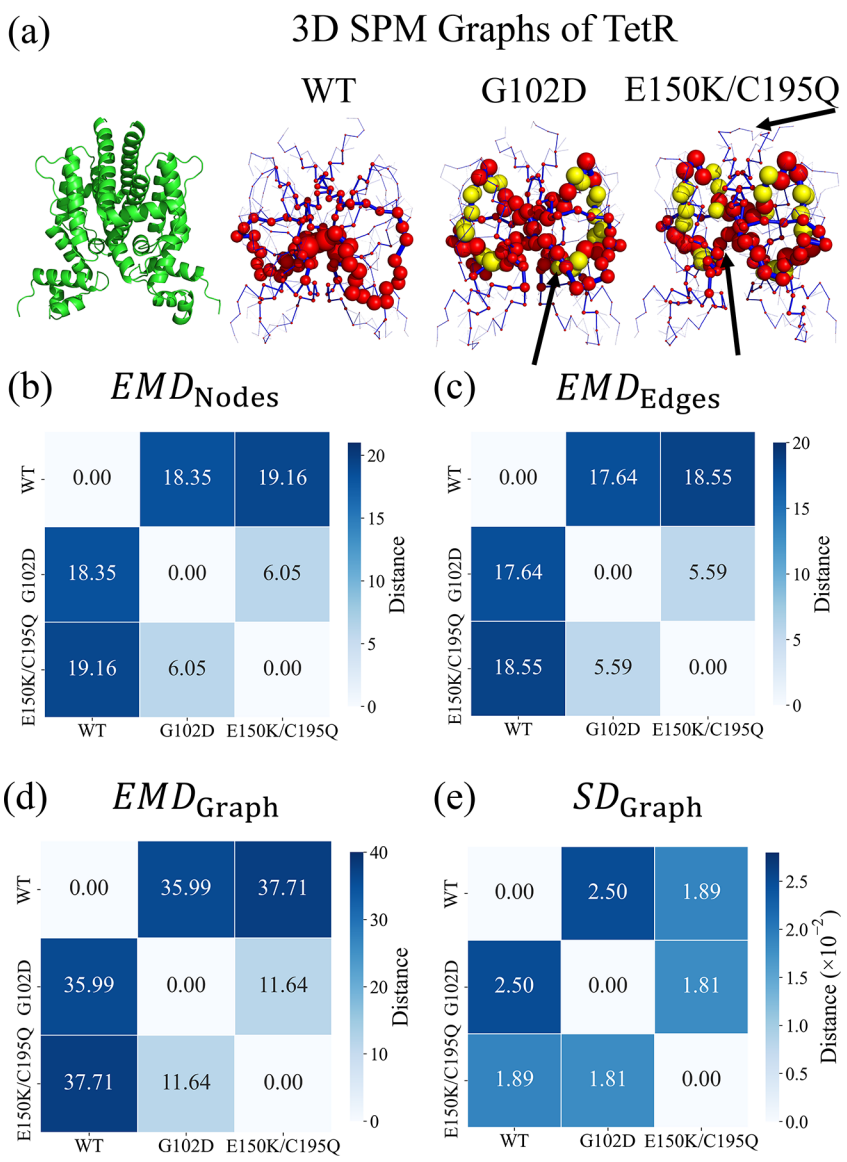



**Figure 10.** (a) Three 3D SPM graphs are generated from equilibrium MD trajectories for WT tetracycline repressor and its two variants (G102D and E150K/C195Q). (b) The pairwise $EMD_{Nodes}$ matrix for WT ubiquitin and its six variants. The location of the mutation is indicated by the arrow. The top ten contribution residues for $EMD_{Nodes}$ with WT are marked in yellow (also see Table 5). (c) The pairwise $EMD_{Edges}$ matrix. (d) The pairwise $EMD_{Graph}$ matrix. (e) The pairwise $SD_{Graph}$ matrix.

in Figure 6b−e indicate that 3D SPM graphs are more sensitive to decreases in the distance threshold than to increases. As shown in Table 2, when $t_{thresh}$ further increases from 6 to 9, the total number of edges grows, but the number of edges with scores >0.1 does not increase, and the numbers of edges with scores >0.2 and >0.3 even decrease. The detailed distributions of node scores and edge scores are provided in the Supporting Information (Figures S14−S19). This indicates that further increasing $t_{thresh}$ mainly generates a large number of weak edges with scores <0.1, while at the same time dispersing the scores that should be concentrated along the main communication pathways. Therefore, a $t_{thresh}$value around 6 is an optimal choice for generating 3D SPM graphs, as it distinguishes edge scores and highlights the major routes without producing a massive number of low-score edges.

### 3.4. Independent Replicate Simulations of Ubiquitin

As shown in Figure 7a, three 3D SPM graphs are generated from three independent 1 $\mu$s equilibrium MD trajectories (Traj-1, Traj-2, and Traj-3) of wild-type ubiquitin, all computed with the CHARMM36m force field but starting from different initial configurations. The large nodes and thick edges remain largely consistent, whereas the small nodes and thin edges exhibit variability. The pairwise distance measures $EMD_{Nodes}$, $EMD_{Edges}$, $EMD_{Graph}$, $SD_{Graph}$ in Figure 7b−d quantify variability arising from different initial conditions within the same system due to the MD simulation fluctuations. As shown in these panels, node- and edge-based metrics display similar magnitudes of contribution. The fluctuations in $EMD_{Nodes}$ and $EMD_{Edges}$ are on the level of around 1, the fluctuation in $EMD_{Graph}$ is on the level of around 2, and the fluctuation in $SD_{Graph}$ is approximately $1.2 \times 10^{-2}$. Here, the analysis and comparison of three independent equilibrium MD trajectories starting from different initial configurations provides a quantitative fluctuation baseline for distinguishing signal from the replicate-level variation.


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

### 3.5. Different Force Fields of Ubiquitin

Next, three 3D SPM graphs are generated from equilibrium MD trajectories simulated with three different force fields (CHARMM36m,[78] CHARMM36[79] and AMBER-ff19SB[80]), as shown in Figure 8a. The most populated conformation differs slightly across force fields. More importantly, the force field also changes the underlying distance and correlation matrices, which further alters the structures of 3D SPM graphs. In general, larger differences between force fields lead to larger discrepancies between the corresponding 3D SPM graphs. The pairwise distances between SPMs obtained with different force fields (Figure 8b−d) exceed the replicate-level variation observed in the three replicate WT trajectories with the CHARMM36m force field (Figure 7b−d). Additionally, the distances show that the 3D SPM graph generated with CHARMM36m,[78] is closer to that generated with CHARMM36,[79] compared to that generated with AMBER-ff19SB.[80] This result is consistent with the physical intuition that CHARMM36m[78] and CHARMM36[79] are closely related to each other than to AMBER-ff19SB.[80] Overall, these results indicate that the allosteric communication pathways revealed by SPM are potentially sensitive to force-field choice, underscoring the need for a careful selection of force fields and, whenever possible, cross-validation of SPM analyses using different force field models to identify the most robust trend.

### 3.6. Different Mutations of Ubiquitin

As shown in Figure 9a, seven 3D SPM graphs are generated from equilibrium MD trajectories for WT ubiquitin and its six variants (I3A, L15S, K33G, P38Q, E51D, and V70N). The pairwise distance matrices of WT ubiquitin and its six variants are shown in Figure 9b−d. For K33G and P38Q, their 3D SPM graphs are similar to that of WT from the visualization and all the four types of pairwise distances ($EMD_{Nodes}$, $EMD_{Edges}$, $EMD_{Graph}$, $SD_{Graph}$) to the WT is close to or within the replicate-level variation observed in the three replicate WT trajectories with the CHARMM36m force field (Figure 7b−d). For L15S and V70N, their 3D SPM graphs are markedly different from the WT, and all the four types of pairwise distances with WT clearly exceed the replicate-level variation, indicating substantial differences from the WT in both spatial and spectral domains. For I3A and E51D, their $EMD_{Nodes}$, $EMD_{Edges}$, and $EMD_{Graph}$ with respect to the WT clearly exceed the replicate-level variation, whereas their $SD_{Graph}$ is slightly above the replicate-level variation, indicating that these variants differ from the WT primarily in spatial organization, while their spectral properties remain relatively similar. Together, these results show that our approach quantitatively distinguishes mutations that minimally perturb allosteric communication from those that substantially rewire it, in both spatial and spectral terms.

To localize the major drivers of SPM differences induced by mutations in ubiquitin, $EMD_{Nodes}$ is decomposed into per-node contributions and $EMD_{Edges}$ is decomposed into per-edge contributions, as shown in Table 3 and Table 4. For example, the decomposition of $EMD_{Nodes}$ between WT and E51D shows that the top three contributors in E51D are LEU43, ASP52 and LEU50. An analogous decomposition for edges highlights pathway-level changes: in the $EMD_{Edges}$ analysis of WT-E51D, the dominant edges in E51D are LEU50-ASP51, LEU43-LEU50, and LEU73-ARG74. In this case, the dominant node and edge contributions are located near the mutation site. In contrast, for other variants the dominant contributions can be distal to the mutation. For example, the decomposition of $EMD_{Nodes}$ between WT and L15S, shows that the top three contributors in L15S are LEU50, GLU51 and LEU43; in the corresponding $EMD_{Edges}$ analysis of WT-L15S, the dominant edges in L15S are LEU50-GLU51, LEU73-ARG74, and ARG72-LEU73. These rankings pinpoint the residues and connections most responsible for mutant-induced SPM shifts and indicate whether these shifts are localized near or distant from the mutation site, providing concrete hypotheses and guidance for analyzing allosteric pathway changes with experimental studies.

### 3.7. Different Mutations of Tetracycline Repressor (TetR)

To illustrate the applicability of the distance methods to larger proteins, we included an additional example on the bacterial tetracycline repressor (TetR). TetR is a dimeric allosteric protein with 207 residues per chain, making it substantially larger than ubiquitin and giving rise to a graph with different size and density. Because TetR is a homodimer composed of two chains, the pairwise distance was calculated separately for each corresponding chain and then summed to obtain the overall pairwise distance between TetR WT and the G102D and E150K/C195Q mutants.

As shown in Figure 10, our distance analysis shows both G102D and E150K/C195Q are substantially separated from WT in the spatial metric, indicating that neither mutant remains close to WT in absolute geometric organization, aligning with the qualitative visualization. Although WT-G102D is slightly less distant than WT-E150K/C195Q in this metric, the difference is small, suggesting only a weak relative advantage for G102D at the spatial level. In contrast, the spectral metric provides much stronger separation, with G102D lying markedly farther from WT than E150K/C195Q. Consistent with these graph distance comparisons, prior TetR simulations identified G102D as a severe domain-interface/l6 hotspot mutant that markedly perturbs the free energy landscape, stabilizes the doubly bound basin, and is among the most difficult hotspot mutants to rescue.[81,82] E150K is located in the $\alpha$8 hotspot region of TetR, adjacent to H151, which was previously implicated in modulating interdomain coupling; C195Q is located in the C-terminal $\alpha$10 region. The closer spectral distance between E150K/C195Q to WT in our network-distance analysis is consistent with a combined $\alpha$8/$\alpha$10 perturbation that is weaker than the severe domain-interface effect of G102D.

To localize the major drivers of SPM differences induced by mutations in TetR, $EMD_{Nodes}$ is decomposed into per-node contributions and $EMD_{Edges}$ is decomposed into per-edge contributions, as shown in Table 5 and Table 6. The top ten residue nodes highlighted in yellow in Figure 10a coincide with the $\alpha$6, $\alpha$7, $\alpha$8 and l6 regions that show the clearest visual differences, where the corresponding nodes are small in WT but become enlarged in G102D and E150K/C195Q. This agreement underscores the power of the distance calculation method to convert visually apparent perturbations into a quantitative and interpretable ranking of mutation-affected regions. Furthermore, for G102D, the mutation-site residue 102 is the second-highest contributing residue node, whereas for E150K/C195Q, the closest contributing residue node to either mutation site among the top ten is residue 138, which ranks sixth and is 12 residues away from mutation site 150. Therefore, for G102D, the dominant node and edge contributions are concentrated near the mutation site. In contrast, for E150K/C195Q, the dominant contributions can be distal to the mutation sites.


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

**Table 5. Top Ten Contribution Residue Nodes in $EMD_{Nodes}$ Calculation for TetR Mutation**

| | WT-G102D | | WT-E150K/C195Q | |
|---|---|---|---|---|
| node | WT | G102D | WT | E150K/C195Q |
| 1 | GLU157 | GLN109 | ARG158 | ARG104 |
| | 1.0376 | 0.9666 | 1.2881 | 1.2233 |
| 2 | ARG158 | ASP102 | GLU156 | THR103 |
| | 0.9572 | 0.8322 | 1.2092 | 1.1768 |
| 3 | VAL153 | THR103 | GLU157 | TYR132 |
| | 0.9156 | 0.8216 | 1.0895 | 1.0014 |
| 4 | GLU156 | LEU117 | VAL153 | GLN116 |
| | 0.8151 | 0.8099 | 1.0691 | 0.9793 |
| 5 | GLN152 | ASN129 | GLN152 | LEU113 |
| | 0.7869 | 0.7967 | 0.9983 | 0.9472 |
| 6 | ASN18 | GLN116 | THR160 | GLY138 |
| | 0.6635 | 0.7868 | 0.9675 | 0.9086 |
| 7 | ALA97 | ARG104 | LYS155 | ASN129 |
| | 0.5926 | 0.7008 | 0.8537 | 0.8999 |
| 8 | LEU146 | ALA130 | LEU146 | GLY102 |
| | 0.5707 | 0.6824 | 0.6705 | 0.8055 |
| 9 | ARG94 | TYR110 | ALA97 | TYR110 |
| | 0.5683 | 0.6732 | 0.6330 | 0.7656 |
| 10 | THR163 | LEU113 | GLU159 | LEU117 |
| | 0.5607 | 0.6138 | 0.4475 | 0.7173 |

These results further highlight the distributed nature of allostery hotspots and their distinct mutational impacts in TetR.[45,81,90,91]

## 4. CONCLUSION

This work developed an algorithm that uses the earth mover's distance (EMD) and the normalized graph Laplacian to quantify pairwise distances between 3D SPM graphs in both spatial and spectral domains. The method jointly evaluates the spatial node and edge ensembles as well as graph connectivity. This approach provides quantitative measures for 3D SPM graphs generated across varying sampling time steps and SPM parameter settings, thereby offering practical guidance for constructing 3D SPM graphs for representing allosteric communication networks. Pairwise distances across independent replicate simulations define a replicate-level variation against which methodological choices can be evaluated. Distances arising from force-field variations exceed this replicate-level variation, indicating that allosteric communication networks generated by molecular dynamics simulations can be sensitive to the choice of force field. Furthermore, the approach quantitatively distinguishes mutations that minimally perturb allosteric communication from those that substantially rewire it. Importantly, decomposition of the EMD highlights residues and edges that undergo the most significant shifts, enabling mechanistic interpretation and informing experimental analyses.[92−94] Apart from illustrating the distance methods in ubiquitin, we also applied them to TetR to demonstrate their generalizability and applicability to graphs that differ substantially from those of ubiquitin.

Together, these results establish a rigorous, extensible basis for benchmarking 3D SPM graphs and quantifying mutation- and model-induced changes in protein allosteric communication networks and conformational dynamics, which is particularly important when a large number of protein variants are compared. Therefore, our approach will greatly facilitate the analysis of high-throughput experiments on enzyme mutation and protein directed evolution.[92−98] More broadly, allosteric coupling is not always reflected in purely concerted correlations, but may also arise through a stepwise sequence of local structural rearrangements.[44,99−101] This suggests that, although 3D SPM graphs provide a useful representation of allosteric communication, alternative graph constructions based on time-lagged or other dynamical measures may be required in some systems. Extending the present distance framework to such graph representations will be an important direction for future work. Furthermore, we note that the approach of quantitatively comparing graphs is not limited to the analysis of protein motions. For example, kinetic properties of biomolecules are well represented by kinetic networks following Markov State

**Table 6. Top Ten Contribution Edges in $EMD_{Edges}$ Calculation for TetR Mutation**

| | WT-G102D | | WT-E150K/C195Q | |
|---|---|---|---|---|
| edge | WT | G102D | WT | E150K/C195Q |
| 1 | ARG158-GLU157 | LEU117-LEU120 | GLN152-VAL153 | ARG104-THR103 |
| | 1.0565 | 0.7816 | 1.3752 | 1.1413 |
| 2 | ALA97-LEU146 | ASP102-LEU101 | ARG158-GLU157 | LEU117-LEU120 |
| | 0.9776 | 0.7664 | 1.1737 | 1.0852 |
| 3 | GLN152-VAL153 | GLN109-THR106 | ALA97-LEU146 | ASN129-TYR132 |
| | 0.9156 | 0.7383 | 1.0485 | 0.9343 |
| 4 | GLU156-LYS155 | LEU113-TYR110 | GLU156-GLU157 | GLN116-LEU113 |
| | 0.6997 | 0.7150 | 0.9799 | 0.9317 |
| 5 | GLU156-GLU157 | GLN109-TYR110 | ARG158-GLU159 | GLY102-THR103 |
| | 0.6632 | 0.7028 | 0.8943 | 0.8904 |
| 6 | ALA154-LYS155 | GLN116-LEU113 | GLU156-LYS155 | GLN116-LEU117 |
| | 0.6312 | 0.6951 | 0.8278 | 0.8276 |
| 7 | ARG94-ASN18 | GLN116-LEU117 | ALA154-VAL153 | ARG104-PRO105 |
| | 0.6261 | 0.6913 | 0.7911 | 0.8232 |
| 8 | ALA154-VAL153 | ARG104-THR103 | GLN152-HSD151 | LEU113-TYR110 |
| | 0.5655 | 0.6640 | 0.7236 | 0.7902 |
| 9 | ASP148-GLU147 | ASN129-TYR132 | PRO161-THR160 | GLU107-TYR110 |
| | 0.5193 | 0.6622 | 0.5216 | 0.6889 |
| 10 | ARG158-GLU159 | ASP102-THR103 | GLU159-THR160 | GLY102-VAL99 |
| | 0.4531 | 0.6218 | 0.5135 | 0.6808 |


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Models;[113] gene regulations in the cellular context are characterized by protein interaction networks.[114] Therefore, we envision that much mechanistic insight can be obtained at different scales if quantitative comparisons can be computed for these graphs as well.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Codes, notebooks and data to run and reproduce all the examples and the corresponding results contained in this article are posted at the GitHub repository https://github.com/shijiale0609/Distance_Protein_Dynamics.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.6c00082.

Details of molecular dynamics simulations, node score distributions and edge score distributions of 3D SPM graphs under varying significance thresholds, node score distributions and edge score distributions of 3D SPM graphs under varying distance thresholds, sensitivity of $EMD_{Node}$ to chemical-distance definitions, justification and sensitivity of $\beta_1$ and $\beta_2$ in Eq. 3 (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Qiang Cui** − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States; Department of Physics, Boston University, Boston, Massachusetts 02215, United States; Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, United States;* orcid.org/0000-0001-6214-5211; Email: qiangcui@bu.edu

### Authors

**Jiale Shi** − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States;* orcid.org/0000-0002-5447-3925

**Zhongyi Wan** − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

**Renjie Zhu** − *Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.6c00082


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The work is supported by the NIH Grant R35-GM141930 to Q.C. We express our gratitude to Professor Osuna, the Osuna Lab and the SPMweb team, who provided the SPMweb API key and guidance to the generation of 3D SPM graphs. Computational resources were provided by Boston University's Shared Computing Cluster, which is administered by Boston University's Research Computing Services (www.bu.edu/tech/support/research/).

## ■ REFERENCES


(1) Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116* (11), 6503−6515.

(2) Cui, Q.; Hamm, P.; Haran, G.; Hyeon, C. Introduction to New Views of Allostery. *J. Chem. Phys.* **2024**, *161* (15), 150401.

(3) Introduction to Protein Ensembles and Allostery. *Chem. Rev.* **2016**, *116* (11), 6263−6266. DOI: .

(4) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. Dynamically Driven Protein Allostery. *Nat. Struct. Mol. Biol.* **2006**, *13* (9), 831−838.

(5) Changeux, J.-P. Allostery and the Monod-Wyman-Changeux Model after 50 Years. *Annu. Rev. Biophys.* **2012**, *41*, 103−133.

(6) Eaton, W. A. A Retrospective on Statistical Mechanical Models for Hemoglobin Allostery. *J. Chem. Phys.* **2022**, *157* (18), 184104.

(7) Konovalov, K. A.; Wu, C.-G.; Qiu, Y.; Balakrishnan, V. K.; Parihar, P. S.; O'Connor, M. S.; Xing, Y.; Huang, X. Disease Mutations and Phosphorylation Alter the Allosteric Pathways Involved in Autoinhibition of Protein Phosphatase 2a. *J. Chem. Phys.* **2023**, *158* (21), 215101.

(8) Tzeng, S.-R.; Kalodimos, C. G. Protein Dynamics and Allostery: An Nmr View. *Curr. Opin. Struct. Biol.* **2011**, *21* (1), 62−67.

(9) Gampp, O.; Kadavath, H.; Riek, R. Nmr Tools to Detect Protein Allostery. *Curr. Opin. Struct. Biol.* **2024**, *86*, 102792.

(10) Grutsch, S.; Brüschweiler, S.; Tollinger, M. Nmr Methods to Study Dynamic Allostery. *PLoS Comput. Biol.* **2016**, *12* (3), No. e1004620.

(11) Feher, V. A.; Durrant, J. D.; Van Wart, A. T.; Amaro, R. E. Computational Approaches to Mapping Allosteric Pathways. *Curr. Opin. Struct. Biol.* **2014**, *25*, 98−103.

(12) Ramanoudjame, G.; Du, M.; Mankiewicz, K. A.; Jayaraman, V. Allosteric Mechanism in Ampa Receptors: A Fret-Based Investigation of Conformational Changes. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (27), 10473−10478.

(13) Fournier, L.; Guarnera, E.; Kolmar, H.; Becker, S. Allosteric Antibodies: A Novel Paradigm in Drug Discovery. *Trends Pharmacol. Sci.* **2025**, *46* (4), 311−323.

(14) Lu, S.; Shen, Q.; Zhang, J. Allosteric Methods and Their Applications: Facilitating the Discovery of Allosteric Drugs and the Investigation of Allosteric Mechanisms. *Acc. Chem. Res.* **2019**, *52* (2), 492−500.

(15) Nussinov, R.; Tsai, C.-J.; Liu, J. Principles of Allosteric Interactions in Cell Signaling. *J. Am. Chem. Soc.* **2014**, *136* (51), 17692−17701.

(16) Wenthur, C. J.; Gentry, P. R.; Mathews, T. P.; Lindsley, C. W. Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165−184.

(17) Changeux, J.-P.; Edelstein, S. J. Allosteric Mechanisms of Signal Transduction. *Science* **2005**, *308* (5727), 1424−1428.

(18) Changeux, J.-P.; Christopoulos, A. Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation. *Cell* **2016**, *166* (5), 1084−1102.

(19) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331−339.

(20) Cui, Q.; Karplus, M. Allostery and Cooperativity Revisited. *Protein Sci.* **2008**, *17* (8), 1295−1307.

(21) Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; Hamm, P.; Stote, R. H.; Eberhardt, J.; Chebaro, Y.; Dejaegere, A.; Cecchini, M.; Changeux, J.-P.; Bolhuis, P. G.; Vreede, J.; Faccioli, P.; Orioli, S.; Ravasio, R.; Yan, L.; Brito, C.; Wyart, M.; Gkeka, P.; Rivalta, I.; Palermo, G.; McCammon, J. A.; Panecka-Hofman, J.; Wade, R. C.; Di Pizio, A.; Niv, M. Y.; Nussinov, R.; Tsai, C.-J.; Jang, H.; Padhorny, D.; Kozakov, D.; McLeish, T. Allostery in Its Many Disguises: From Theory to Applications. *Structure* **2019**, *27* (4), 566−578.

(22) Nussinov, R.; Jang, H. The Value of Protein Allostery in Rational Anticancer Drug Design: An Update. *Expert Opin. Drug Discovery* **2024**, *19* (9), 1071−1085.

(23) Brunori, M. Half a Century of Hemoglobin's Allostery. *Biophys. J.* **2015**, *109* (6), 1077−1079.

(24) Enakaya, N. A.; Jefferson, A.; Chew-Martinez, D.; Matthews, J. S. Design, Synthesis, and Evaluation of Allosteric Effectors for Hemoglobin. *Acc. Chem. Res.* **2023**, *56* (11), 1279−1286.

(25) Perutz, M. F.; Wilkinson, A. J.; Paoli, M.; Dodson, G. G. The Stereochemical Mechanism of the Cooperative Effects in Hemoglobin Revisited. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 1−34.
(26) Yuan, Y.; Tam, M. F.; Simplaceanu, V.; Ho, C. New Look at Hemoglobin Allostery. *Chem. Rev.* **2015**, *115* (4), 1702−1724.
(27) Guo, Z.; Smutok, O.; Lee, G. R.; Cui, Z.; Qianzhu, H.; Kish, M.; Ergun yva, C.; Wu, K.; Mutschler, R.; Jackson, C. J.; Fiorito, M. M.; Warden, A. C.; Smith, O. B.; Quijano-Rubio, A.; Huber, T.; Phillips, J. J.; Otting, G.; Katz, E.; Baker, D.; Alexandrov, K. Artificial Allosteric Protein Switches with Machine-Learning-Designed Receptors. *Nat. Biotechnol.* **2026**.
(28) Helmstaedt, K.; Krappmann, S.; Braus, G. H. Allosteric Regulation of Catalytic Activity:Escherichia Coli Aspartate Transcarbamoylase Versus Yeast Chorismate Mutase. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (3), 404−421.
(29) Hvorecny, K. L.; Hargett, K.; Quispe, J. D.; Kollman, J. M. Human Prps1 Filaments Stabilize Allosteric Sites to Regulate Activity. *Nat. Struct. Mol. Biol.* **2023**, *30* (3), 391−402.
(30) Kremer, D. M.; Lyssiotis, C. A. Targeting Allosteric Regulation of Cancer Metabolism. *Nat. Chem. Biol.* **2022**, *18* (5), 441−450.
(31) Pincus, D.; Pandey, J. P.; Feder, Z. A.; Creixell, P.; Resnekov, O.; Reynolds, K. A. Engineering Allosteric Regulation in Protein Kinases. *Sci. Signal.* **2018**, *11* (555), No. eaar3250.
(32) Kornev, A. P.; Taylor, S. S. Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* **2015**, *40* (11), 628−647.
(33) Brognard, J.; Hunter, T. Protein Kinase Signaling Networks in Cancer. *Curr. Opin. Genet. Dev.* **2011**, *21* (1), 4−11.
(34) Garrington, T. P.; Johnson, G. L. Organization and Regulation of Mitogen-Activated Protein Kinase Signaling Pathways. *Curr. Opin. Cell Biol.* **1999**, *11* (2), 211−218.
(35) Zhu, S.; Sridhar, A.; Teng, J.; Howard, R. J.; Lindahl, E.; Hibbs, R. E. Structural and Dynamic Mechanisms of Gabaa Receptor Modulators with Opposing Activities. *Nat. Commun.* **2022**, *13* (1), 4582.
(36) Sigel, E.; Steinmann, M. E. Structure, Function, and Modulation of Gabaa Receptors. *J. Biol. Chem.* **2012**, *287* (48), 40224−40231.
(37) Swamy, M.; Beck-Garcia, K.; Beck-Garcia, E.; Hartl, F. A.; Morath, A.; Yousefi, O. S.; Dopfer, E.; Molnár, E.; Schulze, A. K.; Blanco, R.; Borroto, A.; Martín-Blanco, N.; Alarcon, B.; Höfer, T.; Minguet, S.; Schamel, W. W. A. A Cholesterol-Based Allostery Model of T Cell Receptor Phosphorylation. *Immunity* **2016**, *44* (5), 1091−1101.
(38) Robinson, J. I.; Baxter, E. W.; Owen, R. L.; Thomsen, M.; Tomlinson, D. C.; Waterhouse, M. P.; Win, S. J.; Nettleship, J. E.; Tiede, C.; Foster, R. J.; Owens, R. J.; Fishwick, C. W. G.; Harris, S. A.; Goldman, A.; McPherson, M. J.; Morgan, A. W. Affimer Proteins Inhibit Immune Complex Binding to Fcγriiia with High Specificity through Competitive and Allosteric Modes of Action. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (1), E72−E81.
(39) Corrada, D.; Morra, G.; Colombo, G. Investigating Allostery in Molecular Recognition: Insights from a Computational Study of Multiple Antibody−Antigen Complexes. *J. Phys. Chem. B* **2013**, *117* (2), 535−552.
(40) Thélot, F. A.; Zhang, W.; Song, K.; Xu, C.; Huang, J.; Liao, M. Distinct Allosteric Mechanisms of First-Generation Msba Inhibitors. *Science* **2021**, *374* (6567), 580−585.
(41) Novischi, S. Y. P.; Karoly-Lakatos, A.; Chok, K.; Bonifer, C.; Becker-Baldus, J.; Glaubitz, C. Probing the Allosteric Nbd-Tmd Crosstalk in the Abc Transporter Msba by Solid-State Nmr. *Commun. Biol.* **2024**, *7* (1), 43.
(42) Procko, E.; O'Mara, M. L.; Bennett, W. F. D.; Tieleman, D. P.; Gaudet, R. The Mechanism of Abc Transporters: General Lessons from Structural and Functional Studies of an Antigenic Peptide Transporter. *FASEB J.* **2009**, *23* (5), 1287−1302.
(43) Thirumalai, D.; Hyeon, C.; Zhuravlev, P. I.; Lorimer, G. H. Symmetry, Rigidity, and Allosteric Signaling: From Monomeric Proteins to Molecular Machines. *Chem. Rev.* **2019**, *119* (12), 6788−6821.
(44) Bozovic, O.; Zanobini, C.; Gulzar, A.; Jankovic, B.; Buhrke, D.; Post, M.; Wolf, S.; Stock, G.; Hamm, P. Real-Time Observation of Ligand-Induced Allosteric Transitions in a Pdz Domain. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (42), 26031−26039.
(45) Leander, M.; Yuan, Y.; Meger, A.; Cui, Q.; Raman, S. Functional Plasticity and Evolutionary Adaptation of Allosteric Regulation. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (41), 25445−25454.
(46) Chure, G.; Razo-Mejia, M.; Belliveau, N. M.; Einav, T.; Kaczmarek, Z. A.; Barnes, S. L.; Lewis, M.; Phillips, R. Predictive Shifts in Free Energy Couple Mutations to Their Phenotypic Consequences. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (37), 18275−18284.
(47) Cui, Q. Identification and Understanding of Allostery Hotspots in Proteins: Integration of Deep Mutational Scanning and Multi-Faceted Computational Analyses. *J. Mol. Biol.* **2025**, *437* (20), 168998.
(48) Zhang, L.; Mobbs, J. I.; May, L. T.; Glukhova, A.; Thal, D. M. The Impact of Cryo-Em on Determining Allosteric Modulator-Bound Structures of G Protein-Coupled Receptors. *Curr. Opin. Struct. Biol.* **2023**, *79*, 102560.
(49) Lü, W.; Du, J.; Goehring, A.; Gouaux, E. Cryo-Em Structures of the Triheteromeric Nmda Receptor and Its Allosteric Modulation. *Science* **2017**, *355* (6331), No. eaal3729.
(50) Turner, M.; Harkness, R. W.; Ripstein, Z. A.; Huang, R.; Vahidi, S. Allosteric Regulation of Proteolytic Machines Unveiled by the Synergy between Cryo-Em and Solution Nmr Spectroscopy. *Biochem. J.* **2025**, *482* (17), 1229−1251.
(51) Woods, V. A.; Abzalimov, R. R.; Keedy, D. A. Native Dynamics and Allosteric Responses in Ptp1b Probed by High-Resolution Hdx-Ms. *Protein Sci.* **2024**, *33* (6), No. e5024.
(52) Kielkopf, C. S.; Ghosh, M.; Anand, G. S.; Brown, S. H. J. Hdx-Ms Reveals Orthosteric and Allosteric Changes in Apolipoprotein-D Structural Dynamics Upon Binding of Progesterone. *Protein Sci.* **2019**, *28* (2), 365−374.
(53) Ducharme, J.; Polic, V.; Thibodeaux, C. J.; Auclair, K. Combining Small-Molecule Bioconjugation and Hydrogen−Deuterium Exchange Mass Spectrometry (Hdx-Ms) to Expose Allostery: The Case of Human Cytochrome P450 3a4. *ACS Chem. Biol.* **2021**, *16* (5), 882−890.
(54) Casadevall, G.; Casadevall, J.; Duran, C.; Osuna, S. The Shortest Path Method (Spm) Webserver for Computational Enzyme Design. *Protein Eng. Des. Sel.* **2024**, *37*, gzae005.
(55) Osuna, S. The Challenge of Predicting Distal Active Site Mutations in Computational Enzyme Design. *WIREs Comput. Mol. Sci.* **2021**, *11* (3), No. e1502.
(56) Shortest Path Map Tool. 2025. https://spmosuna.com/ (accessed Dec 15, 2025).
(57) Curado-Carballada, C.; Feixas, F.; Iglesias-Fernández, J.; Osuna, S. Hidden Conformations in Aspergillus Niger Monoamine Oxidase Are Key for Catalytic Efficiency. *Angew. Chem., Int. Ed.* **2019**, *58* (10), 3097−3101.
(58) Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **2017**, *7* (12), 8524−8532.
(59) Kearns, F. L.; Bogetti, A. T.; Calvó-Tusell, C.; Braza, M. K. E.; Casalino, L.; Gramm, A. J.; Braet, S.; Rosenfeld, M. A.; Rajapaksha, H.; Barker, B.; Anand, G.; Ahn, S.-H.; Chong, L. T.; Amaro, R. E. D614g Reshapes Allosteric Networks and Opening Mechanisms of Sars-Cov-2 Spikes. *bioRxiv* **2025**, *2025*.2003.2007.642081. DOI: .
(60) 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.
(61) Duran, C.; Casadevall, G.; Osuna, S. Harnessing Conformational Dynamics in Enzyme Catalysis to Achieve Nature-Like Catalytic Efficiencies: The Shortest Path Map Tool for Computational Enzyme Redesign. *Faraday Discuss.* **2024**, *252* (0), 306−322.
(62) Herter, S.; Medina, F.; Wagschal, S.; Benhaïm, C.; Leipold, F.; Turner, N. J. Mapping the Substrate Scope of Monoamine Oxidase (Mao-N) as a Synthetic Tool for the Enantioselective Synthesis of Chiral Amines. *Biorg. Med. Chem.* **2018**, *26* (7), 1338−1346.
(63) Buller, A. R.; Brinkmann-Chen, S.; Romney, D. K.; Herger, M.; Murciano-Calles, J.; Arnold, F. H. Directed Evolution of the


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Tryptophan Synthase B-Subunit for Stand-Alone Function Recapitulates Allosteric Activation. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (47), 14599−14604.
(64) Dunn, M. F. Allosteric Regulation of Substrate Channeling and Catalysis in the Tryptophan Synthase Bienzyme Complex. *Arch. Biochem. Biophys.* **2012**, *519* (2), 154−166.
(65) Csárdi, G.; Nepusz, T. *The Igraph Software Package for Complex Network Research*, 2006.
(66) Hargreaves, C. J.; Dyer, M. S.; Gaultois, M. W.; Kurlin, V. A.; Rosseinsky, M. J. The Earth Mover's Distance as a Metric for the Space of Inorganic Compositions. *Chem. Mater.* **2020**, *32* (24), 10610−10620.
(67) Durdy, S.; Hargreaves, C. J.; Dennison, M.; Wagg, B.; Moran, M.; Newnham, J. A.; Gaultois, M. W.; Rosseinsky, M. J.; Dyer, M. S. The Liverpool Materials Discovery Server: A Suite of Computational Tools for the Collaborative Discovery of Materials. *Digit. Discovery* **2023**, *2* (5), 1601−1611.
(68) Shi, J.; Walsh, D.; Zou, W.; Rebello, N.; Deagen, M.; Fransen, K.; Gao, X.; Olsen, B.; Audus, D. Calculating Pairwise Similarity of Polymer Ensembles Via Earth Mover's Distance. *ACS Polym. Au* **2024**, *4*, 66.
(69) Shi, J.; Rebello, N. J.; Walsh, D.; Zou, W.; Deagen, M. E.; Leão, B. S.; Audus, D. J.; Olsen, B. D. Quantifying Pairwise Similarity for Complex Polymers. *Macromolecules* **2023**, *56*, 7344.
(70) Weininger, D.; Weininger, A.; Weininger, J. L. Smiles. 2. Algorithm for Generation of Unique Smiles Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (2), 97−101.
(71) Weininger, D. Smiles, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28* (1), 31−36.
(72) Flamary, R.; Courty, N.; Gramfort, A.; Alaya, M. Z.; Boisbunon, A.; Chambon, S.; Chapel, L.; Corenflos, A.; Fatras, K.; Fournier, N.; Gautheron, L.; Gayraud, N. T. H.; Janati, H.; Rakotomamonjy, A.; Redko, I.; Rolet, A.; Schutz, A.; Seguy, V.; Sutherland, D. J.; Tavenard, R.; Tong, A.; Vayer, T. Pot: Python Optimal Transport. *J. Mach. Learn. Res.* **2021**, *22* (78), 1−8.
(73) Flamary, R.; Vincent-Cuaz, C.; Courty, N.; Gramfort, A.; Kachaiev, O.; Quang Tran, H.; David, L.; Bonet, C.; Cassereau, N.; Gnassounou, T.; Tanguy, E.; Delon, J.; Collas, A.; Mazelet, S.; Chapel, L.; Kerdoncuff, T.; Yu, X.; Feickert, M.; Krzakala, P.; Liu, T.; Fernandes Montesuma, E. *Pot Python Optimal Transport* (Version 0.9.5). 2024. https://github.com/PythonOT/POT (accessed Dec 15, 2025).
(74) Chen, H.; Zhang, F. Resistance Distance and the Normalized Laplacian Spectrum. *Discrete Appl. Math.* **2007**, *155* (5), 654−661.
(75) Bauer, F. Normalized Graph Laplacians for Directed Graphs. *Linear Algebra Appl.* **2012**, *436* (11), 4193−4222.
(76) Wu, X. M.; Li, Z.; Chang, S. F. New Insights into Laplacian Similarity Search. *2015 IEEE Conference on Computer Vision and Pattern Recognition* **2015**, 1949−1957.
(77) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8å Resolution. *J. Mol. Biol.* **1987**, *194* (3), 531−544.
(78) 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.
(79) Huang, J.; MacKerell Jr, A. D. Charmm36 All-Atom Additive Protein Force Field: Validation Based on Comparison to Nmr Data. *J. Comput. Chem.* **2013**, *34* (25), 2135−2145.
(80) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19sb: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528−552.
(81) Yuan, Y.; Deng, J.; Cui, Q. Molecular Dynamics Simulations Establish the Molecular Basis for the Broad Allostery Hotspot Distributions in the Tetracycline Repressor. *J. Am. Chem. Soc.* **2022**, *144* (24), 10870−10887.
(82) Deng, J.; Yuan, Y.; Cui, Q. Modulation of Allostery with Multiple Mechanisms by Hotspot Mutations in Tetr. *J. Am. Chem. Soc.* **2024**, *146* (4), 2757−2768.
(83) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4* (2), 187−217.
(84) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; 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.; Hodoscek, 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. C. The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.
(85) 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.
(86) Park, S.; Choi, Y. K.; Kim, S.; Lee, J.; Im, W. Charmm-Gui Membrane Builder for Lipid Nanoparticles with Ionizable Cationic Lipids and Pegylated Lipids. *J. Chem. Inf. Model.* **2021**, *61* (10), 5192−5202.
(87) Kim, S.; Lee, J.; Jo, S.; Brooks III, C. L.; Lee, H. S.; Im, W. Charmm-Gui Ligand Reader and Modeler for Charmm Force Field Generation of Small Molecules. *J. Comput. Chem.* **2017**, *38* (21), 1879−1886.
(88) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. Charmm-Gui Membrane Builder toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014**, *35* (27), 1997−2004.
(89) 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.
(90) Leander, M.; Liu, Z.; Cui, Q.; Raman, S. Deep Mutational Scanning and Machine Learning Reveal Structural and Molecular Rules Governing Allosteric Hotspots in Homologous Proteins. *eLife* **2022**, *11*, No. e79932.
(91) Liu, Z.; Gillis, T. G.; Raman, S.; Cui, Q. A Parameterized Two-Domain Thermodynamic Model Explains Diverse Mutational Effects on Protein Allostery. *eLife* **2024**, *12*, RP92262.
(92) Hayes, T.; Rao, R.; Akin, H.; Sofroniew, N. J.; Oktay, D.; Lin, Z.; Verkuil, R.; Tran, V. Q.; Deaton, J.; Wiggert, M.; Badkundri, R.; Shafkat, I.; Gong, J.; Derry, A.; Molina, R. S.; Thomas, N.; Khan, Y. A.; Mishra, C.; Kim, C.; Bartie, L. J.; Nemeth, M.; Hsu, P. D.; Sercu, T.; Candido, S.; Rives, A. Simulating 500 Million Years of Evolution with a Language Model. *Science* **2025**, *387* (6736), 850−858.
(93) Xiao, H.; Bao, Z.; Zhao, H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54* (16), 4011−4020.
(94) Shen, D.; Wang, X.; Gao, Y.; Wang, W.; Li, Y.; Chen, H.; Guo, Y.; Cao, S.; Huang, Y.; Zhang, Y.; Wang, C.; Zhang, S. An Industrial Automated Laboratory for Programmable Protein Evolution. *Nat. Chem. Eng.* **2025**, *2*, 685.
(95) Markin, C. J.; Mokhtari, D. A.; Sunden, F.; Appel, M. J.; Akiva, E.; Longwell, S. A.; Sabatti, C.; Herschlag, D.; Fordyce, P. M. Revealing Enzyme Functional Architecture Via High-Throughput Microfluidic Enzyme Kinetics. *Science* **2021**, *373* (6553), No. eabf8761.
(96) Gantz, M.; Neun, S.; Medcalf, E. J.; van Vliet, L. D.; Hollfelder, F. Ultrahigh-Throughput Enzyme Engineering and Discovery in in Vitro Compartments. *Chem. Rev.* **2023**, *123* (9), 5571−5611.
(97) Rix, G.; Williams, R. L.; Hu, V. J.; Spinner, A.; Pisera, A.; Marks, D. S.; Liu, C. C. Continuous Evolution of User-Defined Genes at 1 Million Times the Genomic Mutation Rate. *Science* **2024**, *386* (6722), No. eadm9073.
(98) Cui, Q. Approaches for Regulating Enzyme Activities: Recent Advances in Experiment and Computation. *Curr. Opin. Struct. Biol.* **2025**, *94*, 103124.
(99) Guénon, P.; Chauveau, M.; Filipowska, K.; Tuñón, I.; Nizak, C.; Reynolds, K.; Stirnemann, G.; Rivoire, O.; Laage, D. Mechanism of Catalytic Regulation by an Allosteric Hot Spot in Dihydrofolate Reductase. *ChemRxiv* **2026**, *2026*(0410)..


https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(100) Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the Sars-Cov-2 Main Protease. *JACS Au* **2023**, *3* (6), 1767−1774.
(101) Ma, L.; Cui, Q. Activation Mechanism of a Signaling Protein at Atomic Resolution from Advanced Computations. *J. Am. Chem. Soc.* **2007**, *129* (33), 10261−10268.
(102) Doering, N. P.; Taterra, M.; Bermudez, M.; Wolber, G. MDPath: Unraveling Allosteric Communication Paths of Drug Targets through Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2025**, *65*, 11123.
(103) Yehorova, D.; Di Geronimo, B.; Robinson, M.; Kasson, P. M.; Kamerlin, S. C.L. Using residue interaction networks to understand protein function and evolution and to engineer new proteins. *Current Opinion in Structural Biology* **2024**, *89*, 102922.
(104) Wang, X.; Anderson, R. M.; Liu, J.; Batista, V.; Loria, J. P. Distal Mutations Rewire Allosteric Networks to Control Substrate Specificity in PTP1B. *Biochemistry* **2025**, *64*, 4661.
(105) Gheeraert, A.; Pacini, L.; Batista, V. S.; Vuillon, L.; Lesieur, C.; Rivalta, I. Exploring Allosteric Pathways of a V-Type Enzyme with Dynamical Perturbation Networks. *J. Phys. Chem. B* **2019**, *123*, 3452.
(106) Yao, X.-Q.; Hamelberg, D. Residue−Residue Contact Changes during Functional Processes Define Allosteric Communication Pathways. *J. Chem. Theory Comput.* **2022**, *18*, 1173.
(107) Ali, A. A. A. I.; Dorbath, E.; Stock, G. Allosteric Communication Mediated by Protein Contact Clusters: A Dynamical Model. *J. Chem. Theory Comput.* **2024**, *20*, 10731.
(108) Chan, H. T. H.; Oliveira, A. S. F.; Schofield, C. J.; Mulholland, A. J.; Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *ACS Au* **2023**, *3*, 1767.
(109) Huizing, G.-J.; Peyre, G.; Cantini, L. Optimal transport improves cell−cell similarity inference in single-cell omics data. *Bioinformatics* **2022**, *38*, 2169.
(110) Alvarez-Melis, D.; Fusi, N. Geometric Dataset Distances Via Optimal Transport. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21428−21429.
(111) *Rdkit*. https://www.rdkit.org/ (accessed 2025-15-12).
(112) Bajusz, D.; Racz, A.; Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* **2015**, *7* (1), 20.
(113) Konovalov, K. A.; Unarta, I. C.; Cao, S.; Goonetilleke, E. C.; Huang, X. Markov State Models to Study the Functional Dynamics of Proteins in the Wake of Machine Learning. *JACS Au* **2021**, *1*, 1330.
(114) Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*; Chapman and Hall/CRC: 2006. DOI: 10.1201/9781420011432




https://doi.org/10.1021/acs.jctc.6c00082
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX