Signal Propagation in Proteins and Relation to Equilibrium Fluctuations

Chakra Chennubhotla, Ivet Bahar*

Department of Computational Biology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

Elastic network (EN) models have been widely used in recent years for describing protein dynamics, based on the premise that the motions naturally accessible to native structures are relevant to biological function. We posit that equilibrium motions also determine communication mechanisms inherent to the network architecture. To this end, we explore the stochastics of a discrete-time, discrete-state Markov process of information transfer across the network of residues. We measure the communication abilities of residue pairs in terms of hit and commute times, i.e., the number of steps it takes on an average to send and receive signals. Functionally active residues are found to possess enhanced communication propensities, evidenced by their short hit times. Furthermore, secondary structural elements emerge as efficient mediators of communication. The present findings provide us with insights on the topological basis of communication patterns that are inherent to the native architecture. To explore the validity and implications of this concept, we assume a discrete-time, discrete-state Markov process of “information” transfer across the network of residues and measure two basic quantities: hitting time and commute time. Hitting time \( H(i,j) \) is the expected number of steps it takes to send information from residue \( i \) to residue \( j \), and this may not be the same as \( H(j,i) \). Commute time \( C(i,j) \) is by definition the sum: \( H(i,j) + H(j,i) \). Hitting time has directionality, while commute time does not.

A major goal in this study is to relate the hitting (and commute) times derived from the Markovian stochastics model to the equilibrium fluctuations (mean-square fluctuations and cross-correlations) of residues predicted by EN models, thus bridging the gap between two disciplines, information theory and statistical mechanics. To this end, using the theory of generalized matrix inverses, we show that hitting/commute times can be expressed in terms of the Kirchhoff matrix of inter-residue contacts that underlie the GNM methodology. Additionally, we present new insights into the signal transduction properties of enzymes, the catalytic residues of which are shown to be distinguished by their fast and precise communication abilities.

The paper is organized as follows. The Results are divided into three parts: first we present the Markovian stochastic model of information diffusion developed for exploring the inter-residue communication in proteins. The process is controlled by transition probabilities for the passage/flow of information across the nodes, which in turn is based on the internode affinities derived from atom–atom contacts in the folded structures. Second, we describe the evaluation of hit and commute times, and illustrate these concepts by presenting the application of the methodology to five different enzymes. Strikingly, active residues are distinguished by their effective communication stochastics. Third, we present a rigorous derivation of the mathematical relation between inter-residue hit/commute times, and their fluctuation dynamics derived from purely statistical mechanical theory. This important relation establishes the bridge between information-theoretic quantities evaluated here for proteins and the intrinsic structural dynamics of proteins as described by physics-based models, and provides a new avenue for further examination of protein allostery using the new information-theoretic perspective.

Editor: Michael Levitt, Stanford University, United States of America

Received April 2, 2007; Accepted July 20, 2007; Published September 21, 2007

Copyright: © 2007 Chennubhotla and Bahar. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: EN, elastic network; GNM, Gaussian Network Model

* To whom correspondence should be addressed. E-mail: bahar@ccbb.pitt.edu
Author Summary

In recent years, there has been a surge in the number of studies using network models for understanding biomolecular systems dynamics. Essentially, two different groups of studies have been performed, driven by two different communities. The first is based on molecular biophysics and statistical mechanical concepts. Normal mode analyses using elastic network models lie in this group. The second is based on information theory and spectral graph methods. The present study demonstrates for the first time that signal transduction events directly depend on the fluctuation dynamics of the biomolecular systems, thus establishing the bridge between the (newly proposed) information-theoretic and the (well-established) physically inspired approaches. We have applied the new approach to five different enzymes. Functionally active residues are shown to possess enhanced communication propensities. Furthermore, secondary structural elements emerge as efficient mediators of communication. These results provide us with important insights for protein design and mechanisms of allostery.

Results

Information-Theoretic Description of Network Communication

Affinity matrices. The protein structure is modeled as a network of $n$ nodes, each representative of a given residue $vi$, for $1 < i < n$. The interaction strength, or the affinity $ajp$, between residues $vi$ and $vj$ is defined as

$$aj = \frac{N_{ij}}{\sqrt{N_i N_j}}$$

where $N_{ij}$ is the total number of atom–atom contacts made between residues $vi$ and $vj$ based on a cutoff distance of $r_c = 4$ Å and $(N_i, N_j)$ are the total numbers of heavy atoms in the individual residues $(vi, vj)$. The affinity matrix is very similar to a contact map or the adjacency matrix of a graph. Non-zero entries in the adjacency matrix denote if two residues are in contact, while the entries in the affinity matrix scale with the number of atom–atom contacts between the residue pair. This representation captures to a first approximation the strong (weak) interactions expected to arise between residue pairs with large (small) number of atom–atom contacts. The denominator corrects for the biases induced by size effects, e.g., for the larger number of atom–atom contacts inherently made by larger size amino acids, thus permitting us to assign affinities purely based on differential interactions. Note that a similar expression has been adopted by Brinda and Vishveshwar [15].

The affinities provide a measure of the local interaction density $dj$ at each residue $v_i$ as $dj = \sum_{p} a_{ij}$. Note that we can define an equivalent mass-spring system having stiffness matrix $gI$ where $I$ is defined in terms of the affinity and degree matrices, $A \equiv \{ a_{ij} \}$ and $D = \text{diag}(dj)$ respectively, as

$$gI = D - A$$

and $g$ is a force constant uniform over all springs. $I$ is also called the Kirchhoff matrix or the combinatorial Laplacian in graph theory [16].

Markov model of network communication. A discrete-time, discrete-state Markov process [11,16] is defined by setting the communication probability between residue pairs to be a function of their affinity. In particular, we define

$$m_{ij} = d_j^{-1}a_{ij}$$

as the conditional probability of transmitting information to residue $v_j$ in one time step given that the signal is initially positioned at residue $v_i$. Note, $d_j$ serves as a normalizing factor to ensure

$$\sum_{k=1}^{n} m_{kj} = 1.$$  \[(4)\]

The conditional probability matrix $M = \{m_{ij}\}$, also called the Markov transition matrix, defines the stochastics of information diffusion over the network of residues, via

$$M = AD^{-1}.$$ \[(5)\]

Suppose the probability of initiating the Markov propagation process at node $j$ is $p_j(0)$. Then, the probability of reaching residue $vi$ in one time step is $m_{ij}p_j(0)$. In matrix notation, the probability of ending up on any of the residues $v = [v_1, v_2, ..., v_n]$ after one time step is given by the distribution $p(t) = M p(0)$. Or, after $k$ steps,

$$p(k) = M^k p(0)$$

where $p(k) = [p_1(k), ..., p_n(k)]$ represents the $n$-dimensional vector of the probabilities of residing at node $i$ $(1 \leq i \leq n)$ at step $k$.

Assume there is a path connecting every pair of residues in the network. Then, as the number of steps $\beta$ approaches infinity, $p(\beta)$ approaches a unique stationary distribution $\pi = [\pi_1, \pi_2, ..., \pi_n]$, the elements of which are given by

$$\pi_i = \frac{d_i}{\sum_{k=1}^{n} d_k}$$

Whereas the evolution of the diffusion process is a function of the starting distribution, the stationary distribution is invariant to the details of initiation. In the continuous time limit [17], the change in probabilities follows the master equation $\frac{dp(t)}{dt} = (M - I) p(t)$ with a solution $p(t) = \exp((M - I) t) p(0)$, where $I$ is an identity matrix of dimension $n \times n$. Here, $M - I$ replaces the transition rate matrix, assuming the time elapsed between successive jumps obeys an exponential probability distribution [17]. Note that (i) in this limit, $t$ replaces $k\delta t$ where $\delta t$ is the mean step size implicitly used in the discrete approximation, and (ii) the stationary distribution corresponding to the continuous process is identical to that obtained for the discrete process such that the detailed balance $m_{ij}\pi_j = m_{ji}\pi_i$ holds.

Hit/commute times as a function of Markov transition probabilities. The hitting time $H(i,j)$ is the average number of steps it takes for the information residing at residue $vi$ to be transmitted to residue $vj$ for the first time. We will term residue $v_i$ a broadcaster and residue $v_j$ a receiver.

The calculation of $H(i,j)$ requires the consideration of all possible pathways on the network, each being weighted by the product of transition probabilities along the path, starting from $v_i$ and ending at $v_j$. An efficient recursive formula can be derived for the calculation of $H(i,j)$ as follows. Suppose the passage from $v_i$ to $v_j$ is performed in two stages, from $v_i$ to a neighbor $v_k$ that is one step away, succeeded by probabilistic passages from $v_k$ to the final destination $v_j$. Furthermore, assume we know the hit time $H(i,k)$ from the intermediate
node \( v_h \) to the destination node \( v_j \). Summing over all the intermediate nodes, the hitting time from \( v_i \) to \( v_j \) is simply

\[
H(j, i) = \sum_{k=1}^{n} [1 + H(j, k)]m_{ki} = \sum_{k=1}^{n} m_{ki} + \sum_{k=1, k \neq j}^{n} H(j, k)m_{ki}
\]

(8)

where Equation 4 is used on the first term on the right hand side. By definition, \( H(i, i) = 0 \). Equation 8 provides a self-consistent method for evaluating the hitting time between any two nodes.

The commute time is defined by the sum of the hitting times in both directions, i.e.,

\[
C(i, j) = H(i, j) + H(j, i) = C(j, i)
\]

(9)

Note that the commute time is symmetric by definition while \( H(i, j) \) is not, as will be illustrated below for example proteins. See the section “Pedagogical example to compute hit/com-mute time” in Methods for hit time analysis of a simple network.

In the calculations below, it proves convenient to define the average hitting times in both directions, as well as the average commute time, for each individual residue as

\[
\langle H_r(j) \rangle = \frac{\sum_i H(j, i)}{n}
\]

\[
\langle H_b(i) \rangle = \frac{\sum_j H(j, i)}{n}
\]

(10)

\[
\langle C(i) \rangle = \frac{\sum_j C(i, j)}{n} = \frac{\sum_j C(j, i)}{n}
\]

\(<H_r(j)> \) and \(<H_b(i)> \) provide a measure of the respective receiver and broadcasting properties of residues \( j \) and \( i \), and \(<C(i)> \) provides a measure of signal transduction properties, in general. Commute time is also known by the name of “resistance distance” in computational chemistry [14,18].

**Application to Enzymes**

**Mapping of hit/commute times between all residue pairs.** Figure 1A displays the hitting times \( H(j, i) \) computed for all residue pairs \((v_i, v_j)\) for an example enzyme, snake phospholipase A2 (Protein Data Bank (PDB) [19], 1bk9 [20]). The blue regions correspond to short hit times, and red regions to long hit times, as indicated by the scales on the right. The map consists of the elements of the hitting time matrix \( H \). Accordingly, the \( j^{th} \) row indicates the number of steps required for a signal to hit residue \( v_j \), starting from any residue \( v_i \). The values are relatively uniform within each row, which reveals that no single residue stands out as an efficient broadcaster, i.e., \(<H_b(i)> \geq 340.3 \pm 1.5 \) for all \( i \). On the other hand, comparing different rows we note that some residues are much better receivers than others: \(<H_r(j)> \) values indeed vary over a broader range of 340.3 ± 124.8.

The higher ability of particular residues to transduce signals is also reflected in the commute times displayed in Figure 1B. The commute time is symmetric by definition, but the hitting time is not. The blue regions along the diagonal show that there is efficient communication along sequential residues, although we also observe several sequentially distant residue pairs that efficiently communicate. While the majority of these residue pairs are spatially close, as will be shown below, there is not necessarily a one-to-one correspondence between commute times and spatial distances, and some residue pairs emerge as more efficient communicators than others despite their longer physical separation.

**Communication properties of individual residues.** Figure 1C displays the mean hitting time \(<H_r(j)> \) for each residue. Minima in this curve point to residues that are effective in receiving signals. It is worth mentioning that the mean commute time \(<C(i)> \), which involves both receive and broadcast times, will have the same profile shape as the mean

![Image](image-url)
Fluctuations determine communication. Consider the hitting
time to the $n^{th}$ residue $v_i$, starting from residue $v_j$. Using Equation 8 and substituting $n$ for $j$, we can use a truncated
version of the Markov transition matrix $\hat{M}$ where the $n^{th}$ row
and column are deleted to obtain

$$\hat{H}_n = \hat{I}^T + \hat{H}_n M.$$  

(11)

Here, $\hat{H}_n$ denotes the $n^{th}$ row of the hitting time matrix $H$
truncated to the first $n$ - 1 elements and $\hat{I}^T$ is a row vector
of length $(n - 1)$ of all 1’s. $M$ can be expressed in terms of a
similarly truncated Kirchhoff matrix $\Gamma = \hat{D} - \hat{A}$ using Equations
2 and 5, leading to

$$H(j, i) = \sum_{k=1}^{n-1} \left\{ [\Gamma^{-1}]_{ki} - [\Gamma^{-1}]_{kj} - [\Gamma^{-1}]_{ji} - [\Gamma^{-1}]_{ji} \right\} d_k$$  

(14)

for the hitting time from residue $v_i$ to any arbitrary residue $v_j$.
Substituting from Equation 19 in Methods, where the elements of the inverse of the Kirchhoff matrix are related to residue
fluctuations [25,26], we obtain

$$H(j, i) = \gamma \sum_{k=1}^{n} \left[ \langle \Delta r_i \Delta r_j \rangle - 2 \langle \Delta r_i \Delta r_j \rangle - \langle \Delta r_i \Delta r_j \rangle + \langle \Delta r_i \Delta r_j \rangle \right] d_k.$$  

(15)

The above equation constitutes the most important result
from the present study: it provides the physical basis for the
hitting times obtained with the information-theoretic meth-
odology by relating them to correlations between residue
fluctuations derived from statistical mechanical theory
[25,26]. The meaning of Equation 15 will be further
elaborated below upon assessment of the contribution of
each term in brackets.

Substitution of Equation 14 in Equation 9 yields an
expression for the commute time in terms of $\Gamma^{-1}$,

$$C(i, j) = \left( [\Gamma^{-1}]_{ji} - [\Gamma^{-1}]_{ij} \right) \sum_{k=1}^{n} d_k$$  

(16)

which, using Equation 19, reduces to

$$C(i, j) = \gamma \sum_{k=1}^{n} \left[ \langle \Delta r_i \Delta r_j \rangle - 2 \langle \Delta r_i \Delta r_j \rangle + \langle \Delta r_i \Delta r_j \rangle \right] d_k.$$  

(17)

This is our final expression bridging commute times with
fluctuations $\langle \Delta r_i \Delta r_j \rangle$ in inter-residue distances. Note that
the term in parentheses is a constant for all pairs of residues.
Thus, the commute time between residues $v_i$ and $v_j$ is directly
proportional to the fluctuations in the distance between these
two residues, larger fluctuations entailing longer commute
times, and vice versa.

More on the physical meaning of hitting times. The hitting

---

**Figure 2.** Average Hit Times $<H(j)>$ versus Their Standard Deviations $\sigma(H(j))$

Catalytic residues (red crosses) are fast and precise, being located at
the lower left end of the plot. Ligand-binding residues are indicated by black +.

---

**Figure 3.** Residues with relatively short hitting times and small deviations
are distinguished in each case by its high communication speed and precision.

---

**Figure 4.** Residues with more pronounced variations in their hitting
and communications times are marked as red crosses. We
distinguish in each case by its high communication speed and precision.

---

**Figure 5.** Residues with small deviations in their hitting
and communications times are marked as red crosses. We
distinguish in each case by its high communication speed and precision.
time expression Equation 14 involves three different types of contributions: a one-body term that depends on the destination node, \( c_j \sum_{i=1}^n d_i \); a two-body term that depends on the initial and final nodes, \(-c_j \sum_{i=1}^n d_i\); and a series of three-body terms that depend on intermediate nodes, in addition to the two end points, \( \sum_{i=1}^n (c_j) d_k \). Of interest is to understand the relative contributions of these three terms. Note that the first is always positive, and increases with the size of destination residue fluctuations; the second may be positive or negative, and the negative sign in front of this term implies that positively correlated residue pairs shorten the hitting time. Likewise, the third term may be positive or negative.

Figure 4 shows the results for phospholipase A2. Figure 4A–4C corresponds to the respective one-body, two-body, and three-body contributions. Note that Figure 4A–4C has different scales, for clearer visualization. As we demonstrate in Figure 4A, the one-body term plays by far a dominant role in determining the resulting hitting times (shown in Figure 1A), i.e., the mean-square fluctuations of the destination node largely determine the hitting time. Residues subject to large amplitude fluctuations require a longer time to be hit, while those subject to small amplitude fluctuations, usually confined to the core or high-density regions, display short hitting times.

The two-body term may be positive or negative, depending on the type of cross-correlations between residues \( v_i \) and \( v_j \). The negative sign on the two-body term implies that the hitting time is reduced if a residue pair undergoes positively correlated fluctuations (Figure 4B). Anticorrelated residues \( (c_j < 0) \), on the other hand, make a positive contribution to Equation 14, thus increasing the communication time. Finally, from Figure 4C we observe that the contribution from the three-body term is negligibly small.

The qualitative features observed here were verified to be valid for all examined proteins: mainly, the mean-square fluctuations of the destination node play a dominant role in determining the hitting (or commute) time, and the cross-correlations between the two end points may increase or decrease the hit/commute time, depending on the type of correlation. Anticorrelations have a retarding effect, while positive correlations reduce the hitting time. In the extreme case of the two nodes moving in phase, by the same amplitude, the effective hit/commute time approaches zero.

**Discussion**

**Communication Distances versus Physical Distances**

The commute times provide us with a means of estimating effective communication distances \( s_{ij} \) between residues \( v_i \) and \( v_j \) using the simple relation
The mean-square distance traveled by a random walk of \( n \) steps, with \( l \) being the average step size. In our case, \( l \) can be readily estimated from the average distance between connected nodes in the network. For phospholipase A2, \( l \) is evaluated to be 3.41 Å. Note that this is shorter than the distance (~3.81 Å) between consecutive \( \alpha \)-carbons, because side chain atoms between neighboring residues may get closer to each other. The number of steps, on the other hand, is directly given by the hitting times themselves (as hitting times are expressed in terms of number of steps). For simplicity, we will use \( n = \frac{1}{2} C(i,j) \) for the effective number of steps for communication between residues \( v_i \) and \( v_j \). The average commute time is similar to the recently proposed “diffusion distance” metric for graphs introduced in previous studies [27,28]. The diffusion distance is based on the eigenvalue decomposition of the Markov transition matrix, whereas the hit/commute times are derived from the graph Laplacian.

How Do Effective Communication Distances Correlate with Physical Distances?

Figure 5A displays the results for phospholipase A2. The effective distance \( s_{ij}^{eff} \) (ordinate) is plotted therein against the physical distance \( s_{ij}^{phy} \) directly evaluated from the PDB coordinates, averaged out over all atoms of residues \( v_i \) and \( v_j \). As expected, the effective communication distances increase with physical distance; however, we can also see a broad variability. The points colored red refer to pairs involving the catalytic residue His48.

Figure 5B and 5C displays from two different perspectives, two residues (Trp31 and Ile104) located at the same physical distance (11.35 ± 0.15 Å) from His48, but differing in their
communication distances (69 Å and 46.5 Å) by a factor of approximately 1.5, pointing to the importance of the particular topology, or secondary structural elements, in increasing the effectiveness of communication. The communication with Trp31, located on a loop, turns out to be much slower, in this case. Figure 5D and 5E illustrates the opposite case of two residues (Lys69 and Asn125) that display comparable communication distances (60.5 ± 1.52 Å), while their respective physical distances (12.9 and 21.6 Å) differ by a factor of 1.7, approximately.

Effect of Secondary Structure

The comparison of the effective and actual (physical) communication distances in Figure 5 suggests that secondary structural elements possess higher abilities in processing signals. To test the validity of this conjecture, we analyzed the distributions of hitting times \( H(i,j) \) to residue \( i \), for the three cases where residue \( i \) is a-helical, \( \beta \)-strand, or coiled/disordered. Figure 6 displays the distributions obtained by combining the results for the examined enzymes. Because the average hitting time increases linearly with the size of a given enzyme, the results for each enzyme are normalized with respect to the number of residues \( N \) in each protein, before combining the data. a-helical residues are observed to be the most efficient communicators, succeeded by \( \beta \)-strand residues (intermediate behavior), while the coiled residues are...
slower and exhibit a broader distribution of hitting times. Examination of the individual proteins, on the other hand, reveals that \( \beta \)-strands may exhibit strong dependence on their spatial location in the 3-D structure of the proteins (Figure 7).

As noted above, the mean-square fluctuations of the destination node play a dominant role in determining the hitting (or commute) time. The higher communication propensity of \( x \)-helical, and to some extent \( \beta \)-strand residues may thus be rationalized by the smaller fluctuations of secondary structural elements compared with coiled regions commonly observed in proteins. It is worth noting, however, that these observations hold for single domain proteins. As illustrated in Figure 8 for a multidomain protein, adenylate kinase, the communication between residue pairs belonging to different domains are usually slower than that between pairs in the same domain.

**Connecting Physically Inspired and Information Theoretic Approaches**

Methods based on network models significantly helped in recent years in providing a comprehensible description of the dynamics of biomolecular systems. On the one hand, methods based on fundamental statistical mechanical principles have been proposed for delineating the collective motions of biomolecules [1–9]; on the other, those based on spectral graph theory and machine learning algorithms have been developed for exploring allosteric effects and response to perturbations/mutations in complex systems [29,30]. While these two methodologies concur in their objectives—understanding the complex machinery of biomolecular systems, the connection between these two approaches has been elusive due to their different originating disciplines, as well as the basic quantities they shed light onto: frequency spectrum and normal mode shapes in the former, shortest paths of communication, and hitting/commute times in the latter.

The present study offers a rigorous way of connecting the two approaches, by demonstrating that the commute times between residues \( v_i \) and \( v_j \), derived from Markov propagation formalism, directly scale with the mean-square fluctuations \( \left\langle \Delta r_{ij}^2 \right\rangle \) in inter-residue distances (Equation 17). Alternatively, the hitting times are expressed in terms of the elements of the covariance matrix (Equation 15). The proportionality between commute times/distances and fluctuations in inter-residue distances explains the “intriguing balance” (or high correlation) recently reported between shortest path lengths and residue fluctuations [31,32]. What the hit-time/commute-time metrics show is the predisposition of the network to exhibit particular communication patterns. These metrics point to how information diffuses in the system. For a protein that behaves as an elastic body, these communication patterns are expected to be intimately connected to equilibrium dynamics, and the present approach connects the equilibrium fluctuations to the kinetic perspective of diffusion on the graph. This analysis may be viewed as a first step toward building analytical models for elucidating the pathways of energy flow in complex biomolecular systems, complementing ongoing MD efforts along the same lines [33].

**Biological Implications and Future Prospects**

Notably, the application to example enzymes point to the more efficient communication propensity and precision of catalytic sites (Figures 1–3), to the role of residue fluctuations and their correlations in transmitting information (e.g., delaying effect of anticorrelated pairs), to the structure-encoded differences in the communication abilities of residue pairs, irrespective of their physical distances; and to the importance of both tertiary contact topology and local (secondary) structure in defining effective means of communication (Figures 5 and 6). Also, irrespective of physical distance, interdomain communication tends to be slower than intradomain, as illustrated in Figure 8 for adenylate kinase.

The major advantage of the present stochastic model over the GNM is the fact that the new methodology lends itself to a comprehensive assessment of the communication paths and their efficiency in biomolecular structures. As such, it holds promise for identifying allosteric communication pathways as well as the sites distinguished by high allostERIC potentials, thus providing insights into the design principles of biomolecular machines. The presently observed enhancement in the information transfer properties of catalytic residues and secondary structural elements suggests possible design requirements for efficient enzymatic activity. In this context, it is worth noting the relevant studies by Choe and Sun [34] and Maritan and coworkers [35], which point to the dependence of equilibrium dynamics on secondary structural content/type. It remains to be understood whether such special communication abilities of catalytic residues result from their local packing topology or more global features conferred by evolutionary pressure.

We note that finding suitable experimental setup for probing hit-times is a challenge. In general, the residues/interactions involved in information flow, or the changes in

---

**Figure 8.** Comparison of Effective (Commute) Distances and Physical Distances between Residue Pairs in E. coli Adenylate Kinase (PDB: 4ake) Effective (ordinate) and physical (abscissa) distances between residues in the CORE, LID, and AMPbd domains (see inset), grouped as intradomain and interdomain distances and shown in different colors for each group. Note that communication between residues in the same domain is more efficient than that between residues in two different domains. This is evidenced by the longer commute distance corresponding to interdomain pairing for a given physical distance, compared with that of intradomain pairs. The inset gives a schematic overview of the distance distributions for intradomain and interdomain pairings.

doi:10.1371/journal.pcbi.0030172.g008
inter-residue distances (which directly define the commute times) may be assessed by site-directed mutagenesis and cross-linking experiments as well as spectroscopic methods such as site-directed fluorescence labeling [36] or FRET [37].

Finally, establishing the bridge between these two disciplines will permit us to translate the wealth of concepts and methods developed in information-theoretic approaches, to exploring the signal transduction mechanisms in complex biomolecular systems, thus complementing physically inspired models and methods.

**Methods**

**Positional fluctuations based on the GNM.** Let $\Delta r_i$ and $\Delta r_j$ be the fluctuation vectors from the mean locations of residues $v_i$ and $v_j$. According to the GNM, the cross-correlations between the fluctuations scale as [1]

$$
\langle \Delta r_i^T \Delta r_j \rangle = \frac{3k_BT}{\gamma} [\Gamma^{-1}]_{ij},
$$

(19)

where $k_B$ is the Boltzmann constant and $[\Gamma^{-1}]_{ij}$ denotes the $ij$th element of the inverse Kirchhoff matrix $\Gamma^{-1}$. The individual residue mean-square (ms) fluctuations are obtained by substituting $i = j$ in Equation 3. Equation 19 is based on purely statistical mechanical considerations, originally put forward by Flory and coworkers for polymer networks [25,26].

**Fluctuations in inter-residue distances.** The mean-square fluctuations $\langle \Delta r_i^T \Delta r_j \rangle$ in inter-residue distances $\Delta r_{ij} = \Delta r_j - \Delta r_i$, can be expressed in terms of the fluctuations in position vectors as

$$
\langle \Delta r_i^T \Delta r_j \rangle = \langle \Delta r_i^T \Delta r_i \rangle + \langle \Delta r_i^T \Delta r_j \rangle - 2\langle \Delta r_i^T \Delta r_i \rangle.
$$

(20)

Equation 20 can be rewritten, using Equation 19, in terms of the elements of $\Gamma^{-1}$ as

$$
\langle \Delta r_i^T \Delta r_j \rangle = \frac{3k_BT}{\gamma} \left( [\Gamma^{-1}]_{ii} + [\Gamma^{-1}]_{ij} - 2[\Gamma^{-1}]_{ij} \right).
$$

(21)

By definition, $\Gamma$ is positive semi-definite, i.e., $\Gamma$ has rank $n - 1$, so it cannot be inverted. Instead, its pseudo-inverse is computed after eliminating the contribution from the zero eigenvalue.

**Pedagogical example to compute hit/commute time.** Consider a small undirected network of three nodes connected as in

$$
i \rightarrow j \rightarrow k
$$

Assume that $a_{ij} = a_{jk} = 1$ and hence the degrees $d_i = d_k = 1; d_j = 2$. For a random walk initiated at $i$ (or $k$), it is obvious that it takes just one time step to reach $j$ as $m_0 = m_0 = 1$. So the hitting time $H(j,i) = 1$. Similarly, one might be tempted to conclude that the expected number of time steps to reach $k$ from $i$ is 2. However, this is not the case because once the random walk reaches $j$, there is an equal chance of returning to $i$ or going to $k$, because $m_0 = m_0 = \frac{1}{2}$. This recursive argument can be unrolled in two ways. First, using Equation 8, the hit time $H(k,i)$ can be expressed as

$$
H(k,i) = 1 + H(k,j)m_{ij},
$$

$$
= 1 + H(k,i).
$$

(22)

Similarly,
Substituting for $p$, we obtain
\[ q = \Gamma^{-1} p \]
\[ s = p - \Gamma q \]
and
\[ r^T = \begin{pmatrix} s^{-1} & (1 + q^T q)^{-1} q^T \end{pmatrix} \iff s \neq 0, \]
\[ r^T = \begin{pmatrix} 1 \end{pmatrix} \iff s = 0. \]  
Substituting for $p$ we obtain
\[ q = \Gamma^{-1} p = -\Gamma^{-1} \Gamma 1 = -1, \]
and hence
\[ \Gamma^{-1} = \begin{pmatrix} \Gamma^{-1} + \begin{pmatrix} 1 & r^T \end{pmatrix} \end{pmatrix}. \]

$1 r^T$ is a rank-1 update to $\Gamma^1$ and $r^T$ is a row vector. So, we can tease out $\Gamma^{-1}$ here by
\[ \Gamma^{-1} = \begin{pmatrix} \Gamma^{-1} & 0^T \end{pmatrix} + \begin{pmatrix} 1 \end{pmatrix} r^T \]
such that in component form
\[ \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} = \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{hi} - r_i, \]
\[ = \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{hi} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ro}. \]

Now when we write
\[ \Gamma = \begin{pmatrix} \Gamma^1 \end{pmatrix}, \]
the effect of adding a row vector $1$ to $\Gamma$ is similar to adding the column vector $p$ to $\Gamma$. So, in equation 38, the inverse of $\Gamma^{-1}$ can be expressed as
\[ \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{hi} = \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{hi} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ro}. \]

Putting the inverses in equations 38 and 39 together, we obtain
\[ \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} = \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{hi} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{os}, \]
\[ = \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ks} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ms} + \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ns}. \]
By substituting equation 40 into equation 13, we get
\[ H(n, i) = \sum_{k=n}^{m} \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} d_k. \]
\[ = \sum_{k=n}^{m} \left( \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ks} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ms} + \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ns} \right) d_k. \]

Using symmetry, the summation can be extended to $n$ as in
\[ H(n, i) = \sum_{k=1}^{n} \left( \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ks} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ms} + \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ns} \right) d_k. \]

This derivation of hitting time to $n^{th}$ residue $v_i$ from $v_j$ was for convenience. For any arbitrary residue $v_p$ the hitting time from residue $v_i$ will be
\[ H(j, i) = \sum_{k=1}^{n} \left( \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ki} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ks} - \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ms} + \begin{pmatrix} \Gamma^{-1} \end{pmatrix}_{ns} \right) d_k. \]

Acknowledgments

We gratefully acknowledge the contribution of Sinem Ozcelik to the results on adenylate kinase.

Author contributions. CC and IB conceived and designed the experiments, performed the experiments, analyzed the data, and wrote the paper.

Funding. This study was supported by US National Institutes of Health grant R33 GM068400-01.

Competing interests. The authors have declared that no competing interests exist.
32. Atilgan AR, Turgut D, Atilgan C (2007) Screened non-bonded interactions in native proteins manipulate optimal paths for robust residue communication. Biophys J BioFAST. doi:10.1529/biophysj.106.099440
33. Ishikura T, Yamato T (2006) Energy transfer pathways relevant for long-range intramolecular signaling of photosensory protein revealed by microscopic energy conductivity analysis. Chem Phy Lett 432: 533–537.
34. Choe S, Sun S (2005) The elasticity of α-helices. J Chem Phy 122: 244912.
35. Micheletti C, Lattanzi G, Maritan A (2002) Elastic properties of proteins: Insight on the folding process and evolutionary selection of native structures. J Mol Biol 321: 909–921.
36. Mansoor S, Farrens D (2004) High-throughput protein structural analysis using site-directed fluorescence labeling and the bimane derivative (2-pyridyl)dithiobimane. Biochemistry 43: 9426–9438.
37. Haas E (2005) The study of protein folding and dynamics by determination of intramolecular distance distributions and their fluctuations using ensemble and single-molecule FRET measurements. Chem Phys Chem 6: 858–870.