Intramolecular Communication and Allosteric Sites in Enzymes Unraveled by Time-Dependent Linear Response Theory

Bang-Chieh Huang[1], Chi-Hong Chang-Chein[1] and Lee-Wei Yang[1,2,6,7,*]

[1]*Institute of Bioinformatics and Structural Biology, National Tsing-Hua University, Hsinchu 30013, Taiwan*

[2]*Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Sciences, Academia Sinica, Taipei 11529, Taiwan*

[3]*Department of Life Science, National Tsing Hua University, Hsinchu 30013, Taiwan.*

[4]*Physics Division, National Center for Theoretical Sciences, Hsinchu 30013, Taiwan*

* To whom correspondence should be addressed:
Lee-Wei Yang. Email: lwyang@mx.nthu.edu.tw; Tel: +88635742467; Fax +88635715934

ABSTRACT

It has been an established idea in recent years that protein is a physiochemically connected network. Allostery, understood in this new context, is a manifestation of residue communicating between remote sites in this network, and hence a rising interest to identify functionally relevant communication pathways and the frequent communicators within. Previous studies rationalized the coupling between functional sites and experimentally observed allosteric sites by theoretically discovered high positional/velocity/thermal correlations between these sites. However, for one to systematically discover previously unobserved allosteric sites in any receptor/enzyme providing the position of functional (orthosteric) sites, these high correlations may not be able to identify remote allosteric sites because of a number of false-positives while many of those are located in proximity to the functional site. Also, whether allosteric sites should be found in equilibrium or non-equilibrium state of a protein to be more biologically relevant is not clear, neither is the directionality preference of aforementioned propagating signals. In this study, we devised a time-dependent linear response theory (td-LRT) integrating intrinsic protein dynamics and perturbation forces that excite protein's temporary reconfiguration at the non-equilibrium state, to describe atom-specific time responses as the propagating mechanical signals and discover that the most frequent remote communicators can be important allosteric sites, mutation of which would deteriorate the hydride transfer rate in DHFR by 2 to 3 orders. The preferred directionality of the signal propagation can be inferred from the asymmetric connection matrix (CM), where the coupling strength between a pair of residues is suggested by their communication score (CS) in the CM, which is found consistent with experimentally characterized nonadditivity of double mutants. Also, the intramolecular communication centers (ICCs), having high CSs, are found evolutionarily conserved, suggesting their biological importance.

INTRODUCTION

The protein structures, dynamics and functions have been intensively studied for several decades. One of the long-standing challenges in revealing the structures-dynamics-function relationship is to estimate the synergistic effects of remote sites, such as relatively buried enzyme active site and alternative binding sites for activators/inhibitors – known as the allosteric effect (1-3, 64). Since 60s, pre-existing model (MWC model) (4) and induced fit model (KNF model) (5) have been used to describe such a mechanism driven by noticeable conformational changes upon ligand binding at a non-orthosteric site. However, it has been first proposed in 1984 by Cooper and Dryden (64) and recently confirmed by NMR experiments (6) that remote mutations could impact the activity of functional sites without involving notable conformational changes. One such example is the cyclic adenosine monophosphate (cAMP) binding to the dimeric catabolite activator protein (CAP), a transcriptional activator. Kalodimos's group used the CPMG technique (6) to quantify microsecond-millisecond dynamics in three different liganded states – no ligand, one monomer binding with a ligand, and two bound ligands in the CAP dimer while one in each monomer. It was found that the binding of one ligand actually generates microsecond-millisecond dynamics that is unseen in unliganded and fully liganded states, causing the negative cooperativity in the binding of the second cAMP binding (6).

Another example is the *E.coli* dihydrofolate reductase (DHFR), which catalyzes the reductation of dihydrofolate (DHF), facilitated by the cofactor NADPH, to terahydrofolate (THF) and $NADP^+$ (Fig 1). The catalytic cycles (7, 8), kinetic data (9-13), folding stability (14-16), intermediate structures (17, 18), and binding dynamics (19, 20) and its hydride transfer rate of the mutants (7, 10, 11, 21-32) have been extensively studied. Recently, more evidences establish the dynamical view point of the enzyme catalytic process in the hydride transfer of DHFR (28, 30, 33, 34). One mutant G121V, 16Å away from the catalytic center, causes no noticeable structural changes (8, 20, 33) and can reduce the hydride transfer rate by 200 folds (10, 11, 26) while other remote mutants G67V (11), S148A (7, 23) and W133F (28) impact the activity only minimally. The QM/MM study also suggests that when a reaction progresses from the reactants to products through the transition states of the protein and/or the chemical compound, a network of "coupled promoting motions" could facilitate the chemistry in DHFR (12, 34). Furthermore, a recent NMR CPMG experiment showed that dynamics on the microsecond-millisecond time scales are significantly changed in the G121V mutant of DHFR, owing to the mutant inducing impaired residue network inside DHFR, which decreases dynamical signal propagation (33).

Besides NMR, other spectroscopy data also suggested that low frequency collective vibrational excitations exist in *ps-ns* range could mediate certain degree of allostery (35-38). Yet, it is still an open question as for how proteins, say enzymes, exploit the vibrational excitations through the signal propagation to regulate catalysis of chemical reactions. In the past decade or two, growing number of theoretical methods were proposed to elucidate intramolecular signal propagation pathways (39-48). In general, the methods can be divided into two categories – structural/dynamical equilibrium approaches (42-44, 46, 47) and those of non-equilibrium (39-41, 45, 48, 49). On the equilibrium approaches, Chennubhotla and Bahar used the Markov approach to extract the communicating residues in a GroEL-GroES system (42, 47). Vishveshwara's group applied Floyd-Warshal algorithm to determine shortest paths between residues having high correlation in positional deviations, analyzed from MD trajectories (43). Kong and Karplus used an unsupervised clustering technique to group residues that are highly correlated in a PDZ domain, and the distributions of communicating residues in a given cluster could report signal 'pathway' (44). Despite the good efforts, one major problem of these

equilibrium approaches is that the propagation pathways are fixed and the perturbations/external forces play no roles in the resolved pathways.

Non-equilibrium approaches served as promising alternatives to investigate signal propagations through the network. Ota and Agard have developed a protocol called "anisotropic thermal diffusion (ATD)" to study the kinetic energy propagation using a non-equilibrium molecular dynamics (MD) simulation (41). The ATD starts with cooling down the protein to 10K in vacuum then a specific residue is heat up to 300K. Consequently, the signal propagation pathway from the heated spot can be tracked by the R.M.S.D. changes of atoms deviated from their equilibrated positions as a function of time. It was found that the speed of intramolecular signaling is about 14 Å/ps, which is comparable to the speed of sound in liquids water at room temperature. However, ATD suffers from low signal-to-noise ratio, which makes it hard to trace the complete network through a protein; also the non-native environment is a concern. Sharp and Skinner used "pump-probe molecular dynamics simulations (PPMD)" to study the signal propagation. Selected residues are 'pumped' by oscillatory forces in a period ~10 ps, the 'probes', are found within those that have the highest enrichment in Fourier-transform-derived density of states for the 'pumped' frequency. The phase delay divided by inter-residue distance is interpreted as the speed of the signal, found at 5Å/ps (45). Although the "mutual correlated residues" were identified by PPMD, the time order of signal propagation is absent. Leitner's group combines the non-equilibrium MD and the normal mode relaxation process to study the diffusivity of heat current in the hemoglobin (39). Given an initial position and velocity, the normal mode relaxation method provides the displacement and velocity as a function of time, which serves as the route to estimate the kinetic energy propagation. Notably, Ranganathan group proposed a correlated mutation method called the "statistical coupling analysis (SCA)" to identity the co-evolved residues (46). This approach, requiring many homologous sequences to secure its statistical significance, however cannot distinguish the structural or dynamics relevance with the found co-evolved mutation sites. One example is that the statistical coupling site W133 shows no obvious effect on the fold reduction of hydride transfer rate in DHFR (28). In fact, the good method, after refinement by maximum-entropy modeling, is later used for protein structure prediction from primary sequences (65).

Supplementing the aforementioned theories, earlier we developed a general time-dependent linear response theory (td-LRT) to investigate the time responses of protein reconfiguration subject to external perturbations (e.g. ligand dissociation) (49). The method, combining normal mode analysis (NMA) (50, 51) and damped harmonic oscillators damped in water, solved by Langevin equation (52), has been shown to reproduce the relaxation time of specific sites in carbonmonoxy myoglobin (Mb) measured by time-resolved UV Resonance Raman (UVRR) spectroscopy (36, 49) and time-resolved X-ray crystallography (53). Using td-LRT, we are able to monitor a two-staged relaxation where the slower relaxation ranges from 4.4 to 81.2 ps, while the faster 'early responses', ranging from hundreds of femtoseconds to a few picoseconds, can be best described by the theory when impulse forces are used. Furthermore, we identify several residues in Mb as 'disseminators' that propagate the signals the fastest and are found kinetically important to modulate gas molecule diffusion and rebinding (49, 56). We further did several point mutations and found that their kinetic importance can be reflected by the number of retained disseminators identified from the wild type (49). The quantitative agreement with several experimental data (36, 53, 56) sets an important distinction of our method.

In this article, we apply this td-LRT theory with impulse forces to investigate the dynamic driven allosteric regulation. We propose a new method to track the atomic-level dynamic signal propagation pathways, and consequently the frequent residue communicators as the intra-molecular communication centers (ICCs). The evolutional conservation and functional importance of these ICCs are carefully examined. The coupling between mechanical perturbations induced by chemical modifications on allosteric sites (or sites remote to the active sites) and fold of the reduction of experimentally characterized hydride transfer rates (at the active site) of DHFR is reported. Finally, we argue that finding spatially clustered ICCs with high communication scores in remote residues can be a promising approach to find allosteric sites.

THEORY, ALGORITHM and METHODS

2a. The System preparation

To study the dynamics relating the forward hydride transfer, the Michaelis-complex of DHFR (PDB code: 1RX2) including cofactor NADPH and substrate DHF is used to perform the normal mode analysis. The DHFR structure is formed by the rigid frame consist of an eight-strand β-sheet and four α-helices as well as three flexible loops: Met 20 loop (residues 9-24), F-G loop (residues 116-132) and G-H loop (residues 142-150) (20).The scalable molecular dynamics program, NAMD (Nanoscale Molecular Dynamics), (57) with CHARMM 36 force-field (58) is used to perform the energy minimization of full atom. Besides, the force-field of substrate, folate, is generated from the ParamChem web site (59, 60) based on the CHARMM General Force Field (CGenFF), where a further Q.M. calculation with MP2/6-31g(d) calculation for the folate support that the charges are close to the estimation of the ParamChem. After several million MD steps of energy minimization, the system reaches our criteria such that the mean force of atoms equals to 10^{-5} kcal/mol/ Å. Consequently, the full atom (*N=2615* atoms including all the hydrogen atoms) normal mode analysis (NMA) is carried out by diagonalizing the hessian matrix constructed by the second derivatives of the potential energy with respect to 3*N* mass-weighting Cartesian coordinates. Further detail steps of the NMA can be found in our previous work (49) and its supporting materials.

2b. The time-dependent linear response (td-LRT) theory and the characteristic time

To study a biomolecule perturbed by the impulse forces, we have developed a time dependent linear response theory (*td*-LRT) to describe the relaxation dynamics (49), which addresses dynamic allostery (64). The time progression of the positional changes of atom *i* are given by the following form

$$\langle \Delta\vec{r}_i(t)\rangle_f = \frac{1}{k_B T}\sum \langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t)\rangle_0 \vec{k}_j \qquad \textbf{(1)}$$

, where $\sum\langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t')\rangle_0$ is the velocity-position time-correlation function sampled in the absence of perturbations, $\vec{k}_j$ are the impulse forces applying on atom *j*. Consider a protein is surrounded by a viscous environment, we express $\langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t')\rangle_0$ as superimposed independent harmonic oscillators (50, 51, 61) under solvent damping by solving the Langevin equation (52). The derivation can be found in supporting materials of our previous work (49). Through our previous studies (49), the speed of signal propagation is a function of locations/directions of perturbation forces and the network architecture within a protein. When impulse forces are introduced on atom *j*, atom *i* responds to these perturbations by its temporary positional departure from its equilibrium position as a function of time. In this

process, the atom reaches its maximal deviation at a characteristic time (t_c^i) before the deviation eventually vanishes at long time (**Fig 2a**). We found that t_c^i is independent of the magnitude of $\vec{k}_j$ but varies with the directions and location of it. To examine the robustness of the perturbation-response relationship, we define a force ensemble, $k(\Omega_j) = k(\theta_j, \phi_j)$, acting at atom *j* with the same magnitude pointing toward 133 directions where θ_j is the angle between the direction of force $\vec{k}_j$ and the z-axis spanning from 0° to 90° with a step size 15°, and ϕ_j is the angle between the projection vector of the force $\vec{k}_j$ on the x-y plane and the x-axis spanning from 0° to 345° with the same step size 15°. Therefore, the characteristic times corresponding to a force ensemble $k(\Omega_j)$ are denoted by $t_c^i(\Omega_j)$.

2c. The longest dissemination time (LDT) relating to the active sites

To understand the role of residues in mediating signals, we would like to measure how fast the signal can propagate throughout the entire enzyme from a residue; in other words, to find the longest response time for all possible signals starting from a specific residue. As in the method proposed in our previous work (49), the force ensemble defined in section 2b was applied on the C_α atom of residue *s*. Consequently, a set of characteristic times consist of $133 \times N$ $t_c^i(\Omega_s)$ can be calculated (N is the number of C_αs or residues, which is 159 for DHFR). The longest $t_c^i(\Omega_s)$ in the *i*-th residue among all the 133×159 $t_c^i(\Omega_s)$ values is the longest dissemination time (LDT) for residue *s*. The residues with short LDT are called the "disseminators" (49), suggesting their roles in efficiently broadcasting the signals throughout the protein matrix. We calculate the LDT for each residue in DHFR by perturbing the corresponding C_α atom, also the C6 atom of the NADPH and N5 atom of the DHF. We then compare these calculated disseminators with the active site residues reported in reference (62).

2d. Coarse-Grained Connection Matrix (CGCM) and communication score (CS)

Although td-LRT provides the characteristic time $t_c^i(\theta_j, \phi_j)$ to characterize the time-evolution of signal propagation, we still need a method to trace frequent signal propagation pathways and identify crucial residues along these pathways, especially when the perturbation is introduced in a number of sites along 133 directions. Here, we introduce "the Coarse-Grained Connection Matrix (CGCM)" to record the signal propagations between residues. The idea is illustrated in **Figure S3**. With the CGCM, we can then extract "the intramolecular communication centers (ICCs)" - the residues having high CSs (see below) on popular pathways communicating dynamic signals.

Because the characteristic time $t_c^i(\theta_j, \phi_j)$ provides us the causality of signal propagation, we can trace the signal transduction pathway consist of the donor-acceptor pairs. There are **two criteria** to define whether a donor atom propagates a signal to an acceptor atom, or say, whether two atoms are viewed as "connected" in the context of signal propagation. The atoms in the donor and acceptor residues are denoted as $atom_d$ and $atom_a$, respectively. **First**, given an impulse force exerted on the *j*-th C_α atom, $\vec{k}_j$, the atoms with characteristic times that differ by a small time interval Δt can be viewed as "connected" such that $t_c^{atom_a} - t_c^{atom_d} = \Delta t$, where the chosen Δt is the constant time interval we store/record the time correlation functions, or equivalently, the evolving protein conformations responding to the perturbation. Δt is 0.2 ps in our study. **Figure S3** demonstrates how the connection signal is recorded into a connection matrix. In order to track the physical causality of signal propagation, the "connected" pairs should also meet the **second** criteria - the angle between

the vector *j*-th C_α to $atom_a$ and *j*-th C_α to $atom_d$ should be less than 90 degree. It is possible that a donor atom connects to several acceptor atoms, or several donors connect to an acceptor atom. As a result, in order to quantitatively recognize how frequently a residue participate in the signal propagation, we define a full-atom connected matrix $F(atom_d, atom_a)$ account for the number of the connection that the donor, a_d, connects to the acceptor, a_r

$$F(atom_d, atom_a) = \sum_j \langle \delta(\Delta t - [t_c^{atom_a}(\Omega_j) - t_c^{atom_d}(\Omega_j)]) \rangle_\Omega \quad (2)$$

where *j* runs over all the selected C_α atoms that are perturbed, $\delta(x)$ is a Delta function equal to zero for any nonzero x, and unity when $x = 0$. Ω_j are forces exerted toward 133 directions defined in section 2b. With the full-atom **F** matrix, we further define the residue level CGCM by summing the counts belong to each residue pair such that

$$C(R_d, R_a) = \sum_{atom_d \in R_d} \sum_{atom_a \in R_a} \frac{F(atom_d, atom_a)}{N_d N_a} \quad (3)$$

where R_d and R_a are donor and acceptor residues, respectively; N_d and N_a are the number of atoms in the donor and acceptor residue, respectively.

To model the signals propagation process during the hydride transfer reaction, we perturbed 21 sites including $C\alpha$ atoms within the distance of 7Å from the catalytic center (N5 atom of the DHF) as well as four sites locating at cofactor and substrate (see **Fig S1**). Consequently, 133 evenly distributed impulse forces defined in section 2b are applying at each perturbed site; all signals are summed up to one CGCM with elements of $C(R_d, R_a)$ which are then normalized with the 133 directions. In the CGCM, the elements in or immediately near the diagonal, indicating intra-residue communication or that between neighboring residues in primary sequence, have the largest values. The high $C(R_d, R_a)$ score (Communication Scores or CSs in short) along diagonals is intuitive for their strong covalent binding but less interesting in terms of allostery. We pay our attention to the off-diagonal elements satisfying $|index(R_d) - index(R_a)| > 2$, which provide information on the signal propagation between long range contacts (say, within or between secondary structures). Several hubs having the highest CGCM scores are the residue pairs that frequently communicate dynamic signals in one or multiple pathways.

Directionality: it is worth noting that the pairwise communication in CGCM is asymmetric; that is to say for residue *i* and *j*, $C(R_i, R_j) \neq C(R_j, R_i)$ despite that the two numbers are usually quite close. As a result, the directionality of signal propagation can be said as from residue *i* to *j* if $C(R_i, R_j) > C(R_j, R_i)$ (see Fig S4 for an example).

For a given residue, a unique "**communication score (CS)**" can be assigned as the highest CGCM score among pairs formed by this residue and any other non-consecutive residue in the protein – in other words, residue *i*'s CS is the highest score in either the *i*-th row or the *i*-th column (corresponding to donors or acceptors) of CGCM. The CS of each residue is listed in **Table S1**.

2e. Evolutionary importance of the ICCs

The evolutionary conservation of residues is compared with their CSs. The sequence conservation of residues are calculated by multiple sequence alignment conducted at the

ConSurf web site (63-65), where residues are categorized into 9 conservation levels by "the ConSurf score" from 1 (the most diverse) to 9 (the most conserved). We then exam the distribution of the ConSurf score for residues with CS larger than several thresholds - 1.1, 1.2, 1.3 and 1.4.

2f. The fold reduction of hydride transfer rates in DHFR mutants is presented as the difference of free energy changes

Suppose that the mutation of an amino acid would break and/or rewire the intrinsic dynamics network, then the chemical modifications (through mutagenesis) in frequently communicating sites, identified by physics approaches (such as td-LRT), could impair the function in the active site allosterically. Here, the forward hydride transfer rate, the chemical step extensively characterized by kinetic isotope experiments (7, 10, 11, 21-32), of wild type DHFR is defined as k_{WT} and that of its mutant is defined as k_{mut}. Converting the reaction rates to its activation free energy, we can use the inverse Boltzmann relation to obtain $\Delta G = -RTln(k)$. Consequently, the difference in activation free energy of hydride transfer reaction between the wild type and a mutant can be written as $\Delta\Delta G = -RTln(\Gamma)$, where $\Gamma = k_{WT}/k_{mut}$ that is the fold reduction of the hydride transfer rate due to a single point mutation. RT = 0.6 kcal/mol. As can be seen in **Table 1**, a large Γ value means that the hydride-transfer rate is largely suppressed by the mutation, while Γ near unity means that the reaction rate stays unchanged. The correlation between the CS of each sites and the corresponding averaged change in free energy difference $\langle\Delta\Delta G\rangle$ is reported in this study, where $\langle ... \rangle$ denotes the average over different mutations at the same site. The results are reported in section 3e.

RESULTS

3a. The signal propagation and characteristic time

Movie S1 and Fig. 2(a) show the relaxation motion of atom *i* (the C_α atom of residue G121) as a function of time, where the point impulse force acting on the atom *j*, the N5 atom of the DHF at the catalytic center, toward the z-direction, is denoted by $(\theta_j, \phi_j) = (0,0)$; the characteristic time, $t_c^i(0,0)$, is 2.6 ps. Considering the force ensemble act at the catalytic center, the red circle in Fig. 2(b) shows the average characteristic time of every single atom *i* over the force ensemble, $\tilde{t}_c^i = \langle t_c^i(\theta_j, \phi_j)\rangle_{\theta,\phi}$, and the black bars denote the corresponding standard deviation. It can be seen that the signal starting from the catalytic center can almost go through the entire DHFR within few picoseconds. Further, $\tilde{t}_c^i$ as a function of the distance between atom *i* and *j*, d_{ij}, is shown in **Fig. 2c**. The linear regression of $\tilde{t}_c^i(d_{ij})$ marked by blue line gives a propagation speed of ~580 *m/s* with a correlation of ~0.8. Generally, the data show that the atom *i* with larger $\tilde{t}_c^i$ is farther in distance from the perturbed site and has larger variance. The variances suggest that the speed of signal is not isotropic and could be affected by the intramolecular dynamic network of residues.

3b. The Longest Dissemination Time (LDT) and disseminators

The dynamic network inside a protein also regulates how fast a perturbation could propagate through the entire protein. In this study we perturb the C_α of all the 159 residues in DHFR, also the C6 atom of the NADPH and N5 atom of the DHF. The LDT of four groups in NADPH and three groups in DHF are also noted (**Fig S2**). For each residue *j*, a perturbation force ensemble in 133 different directions is used and then the slowest $t_c^i(\theta_j, \phi_j)$ at a given residue *i* is noted as the longest dissemination time (LDT) of the residue *j*. The residues with

short LDT are called "disseminators" (49). Figure 3 shows the LDT for all the residues. It was found that several active sites (ILE 5, Phe 31, ILE 94, C6 of NADPH and N5 of DHF) are among good disseminators. The results suggest that the active sites locate at the optimal position of network to efficiently broadcast the mechanical signals.

3c. The Coarse-Grained Connection Matrix (CGCM) and communication score (CS)

Following the procedure in section 2d, the CGCM of DHFR can be obtained (**Figure 4a**). There are several hot spots in the off-diagonal elements, especially, the largest two elements are found to locate at C(13,122)=1.52 and C(13, 121)=1.42 - the donor-acceptor pairs of V13-D122 and V13-G121, respectively. Previous experiments have shown that the mutations in D122 and G121 can significant obstruct the hydride transfer rate. Interestingly, as shown in **Table S3**, we decomposed C(13,121) into 21 components by the perturbed sites, and it is found that signals from the perturbed C6 of the pterin group (site 164) and Y100 are almost twice the size of signals from other perturbed sites. The results provide an evident supporting that a strong coupling between the residue G121 and the catalytic center. Consequently, the CSs are determined following section 2d. **Figure 4b** shows the color-coded DHFR structure by **the CSs**, where the lowest and highest CS correspond to dark blue and dark red, respectively. The first few highest CSs can be found at the residue D122 as well as the second highest CS at the residue G121 located in the F-G loop. Some other residues such like M42, L54 and D27 show moderate CS of 0.6-0.7, while G67, S148 and W133 show CS lower or equal to 0.5. With the current method where pairwise communication is not symmetric, we can assign the directionality for the signal flow. For instance, we can see how the signal is propagated from the allosteric site at G121 and D122 to residues near the catalytic center (**Fig S4**).

After collecting mutagenesis data from 35 mutants at 17 sites of which the hydride transfer rates were carefully characterized without blending in kinetics of substrate binding or product release (7, 10, 11, 21-26, 28), the ratio (Γ) of wide-type hydride transfer rate (7, 10, 11, 21-26, 28) to that of a single mutant can be obtained. It is found that the hydride transfer rate of DHFR could drastically fall by 2 or 3 orders (resulting in a Γ value of hundreds or more than a thousand; see Table 1) if a residue with high CS is mutated. It happens even when the residue is remote (>15Å) to the catalytic center (**Table 1, Fig 6**), suggesting that allosteric regulation on activity of a functional site can be imposed by perturbing (chemically herein, by mutagenesis) spatially clustered ICCs (e.g. G121 and D122) with a high CS, if medicinal interest of this study is concerned.

3d. The conservation of residues of DHFR in evolution

We further examined the evolutional importance of the identified ICCs. The number histogram of the ConSurf scores for all residues in DHFR is shown in **Figure 5** (the black line). There are 31 residues with the lowest ConSurf score (that is 1) as well as 57 residues with the highest ConSurf scores (8 or 9) which dominate the population. As examining the histograms of ConSurf scores for residues having a CS >0.9 (84 residues; in blue line), >1.0 (57 residues), >1.1 (30 residues), >1.2 (16 residues) and >1.3 (5 residues; in red line), we found the population gradually shifts toward evolutionarily conserved regions. The finding supports the aspect that dynamically significant residues are also evolutionarily conserved (66).

3e. The exponential dependence between the hydride transfer rate reduction and CS

By collecting experimental data from the literature, as shown in the **Table 1,** we summarized the fold reduction of hydride transfer rate in 35 mutants at 17 sites. The fold of hydride

transfer rate reduction for a point mutation is expressed as $\Gamma = k_{WT}/k_{mut}$. For example, the I14A mutation causes 40 fold of hydride transfer rate reduction. The red and blue highlights indicate active sites and perturbed sites, respectively. $\langle \Delta\Delta G \rangle = \langle \Delta G_{WT} - \Delta G_{mutant} \rangle = \langle -RTln(\Gamma) \rangle$ is the average for a set of difference of activation free energy in the hydride transfer rates between the wild type and its mutants at a given site (average over the rows in **Table 1**), where $\langle ... \rangle$ denotes the average of $\Delta\Delta G$ for mutants at the same site. **Figure 6** reveals there is an appealing correlation (>0.8) between mutants' -$\langle \Delta\Delta G \rangle$ and their communication scores, suggesting that the frequency of dynamic signal propagation through a residue could play an important role in allosteric control over the hydride transfer rate.

DISCUSSIONS

Is communication frequency between ICCs a function of geometric or physiochemical properties of a protein?

Although the correlation between the allosteric sites and the ICCs is established in our study, the question still remains as for how this "communication" property is related to known geometric or physiochemical features in an enzyme. This is not immediately clear at the moment. Our ENM analysis of the DHFR structure showed that the frequent communicators V13 and G121 are not located in the hinge of the slowest mode (suggesting functional importance (67-71)), nor are they located at the peaks of highest frequency modes (indicating the possibility to serve as a local folding core (72)) (Fig S5). The H-bonds formed between V13 and G121 are nowhere standing out as compared to those formed between other pairwise interactions. Thus far we have not found that the communication frequency of ICCs is correlated with other basic geometric or physiochemical properties such as mechanical hinges (68), binding interface (69-71) or high local packing density (72). Yet, they are evolutionarily conserved and regulatory of enzyme catalysis even when situated remotely.

The nonadditive of the changes of free energy difference for double mutant relating to the coupling of two distal residues also revealed by the elements of CGCM

The coupling between two distal residues has been studied by measuring the changes in free energy difference ($\Delta\Delta G = k_B T ln(\Gamma)$ where $\Gamma = k_{WT}/k_{mut}$) for DHFR's hydride transfer rates between two single mutants and a double mutant of the same two residues (11, 28). It was proposed that the degree of coupling of two distal residues can be measured by the non-additivity of coupling energy ΔG_I (11), where the coupling energy $\Delta_3 G_I$ between two residue A and B is $\Delta_3 G_I = \Delta\Delta G_{A,B} - (\Delta\Delta G_A + \Delta\Delta G_B)$. The larger the magnitude of $|\Delta_3 G_I|$, the higher degree of coupling between the two selected residues is. For instance, given Γ for three single mutants M42W($\Gamma = 41$)(11), G121V($\Gamma = 166$)(11) and F125M($\Gamma = 42$)(28), the coupling energies $|\Delta_3 G_I|$ of three double mutants: M42W-G121V ($\Gamma = 7600$), M42W-F125M ($\Gamma = 512$) and G121V-F125M ($\Gamma = 61$)(28) are 0.077, 0.73 and 2.83, respectively, suggesting positions 42 and 121 are relatively independent, while the 121-125 pair are the most coupled among the three pairs. Interestingly, we found that the communication scores [C(42,121) = 0.26 and C(121,42) = 0.24] are indeed the smallest as compared to [C(42,125) = 0.46 and C(125,42) = 0.45] and [C(121,125) = 0.59 and C(125,121) = 0.66] that is the most frequently communicating pair among the three. Our results show the same order of coupling strength G121V-F125M > M42-F125M > M42W-G121V as that derived from experimentally characterized hydride transfer rates, suggesting the count of propagating signals between a pair of residues can be viewed as a measure to quantify functionally relevant residue-residue coupling in proteins.

The robustness of CS

We have tested several combinations of perturbation sites involving including or excluding p6, p7, p24 to p159. Overall, the conservation dependence (Fig 5) barely changed, and the correlation between CS and hydride transfer rate reduction in mutants varied slightly (Fig 5a, 5b). In addition to the local structure dynamics such as the flexibility predicted by NMR (8,20,33) or contact number from the structural information [], the herein reported CSs could provide a quantitative measurement describing "the propensity of signal propagation between residues". According to the evidence that CSs highly correlate with the hydride transfer rate reduction, our analysis support the aspect that the signal propagation regulates (if not facilitates) the enzyme catalysis, providing an alternative theoretical basis (apart from NMR (8,20,33), bioinformatics, MD) for enzymologists' further exploration.

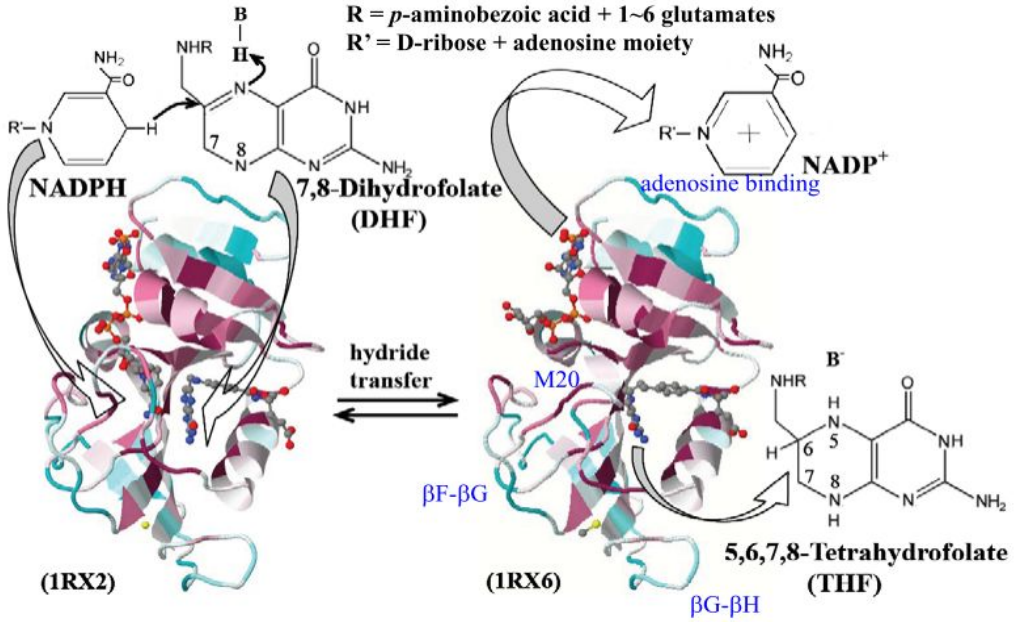

Figure 1 E. coli DHFR ("DHFR" thereafter) facilitates the hydride transfer from NADPH to the C6 of the pterin nucleus with a concurrent protonation in N5 position of DHF in the reaction of NADPH + H^+ + DHF → $NADP^+$ + THF. The hydride transfer-prompting Michaelis (binary) complex on the left (PDB ID: 1RX2) and the product ternary complex on the right (PDB ID:1RX6) are color-coded by residue's evolutionary conservation based on multiple sequence alignment of 77 DHFR homologs from cyan, the most variable, to dark red, the most conserved using Consurf server (63).

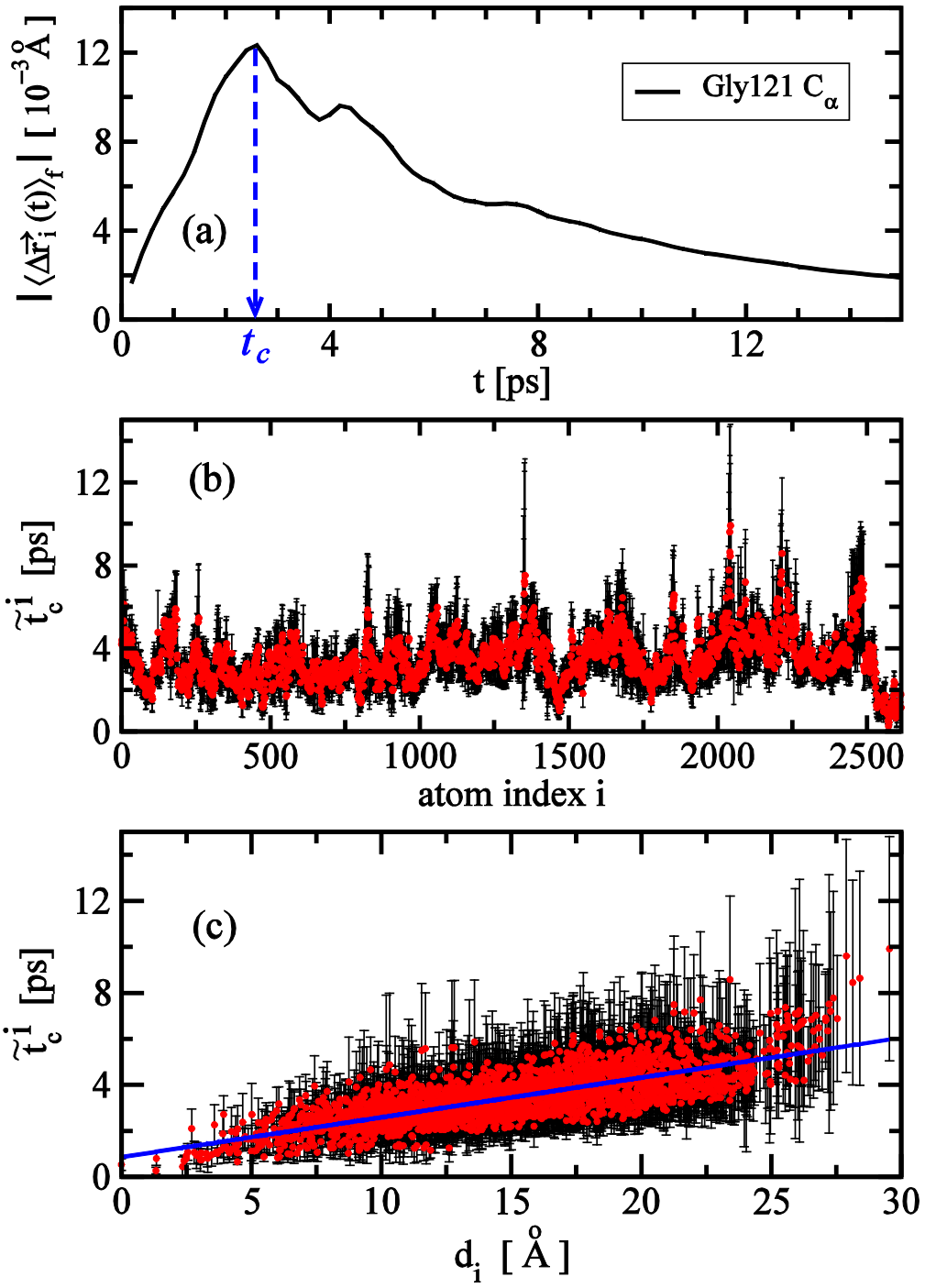

Figure 2 (a) The displacement of Gly 121 C_α atom against the response time with the characteristic time t_c = 2.6 ps predicted by LRT using an impulse force on the N5 atom of the DHF. **(b)** By applying 133 evenly distributed forces on the N5 atom of the DHF, a red dot is the average characteristic times $\tilde{t}_c^i$ of atom *i* and a black error bar is the standard deviations. **(c)** The $\tilde{t}_c^i$ as a function of the distance d_i between atom *i* and the N5 atom of the DHF, where the blue line is the linear regression with the correlation coefficient of 0.76. The inverse slope provided to estimate the propagation speed of 580 *m*/*s*, and the estimated speed is compatible to results from a few theoretical groups (41, 45).

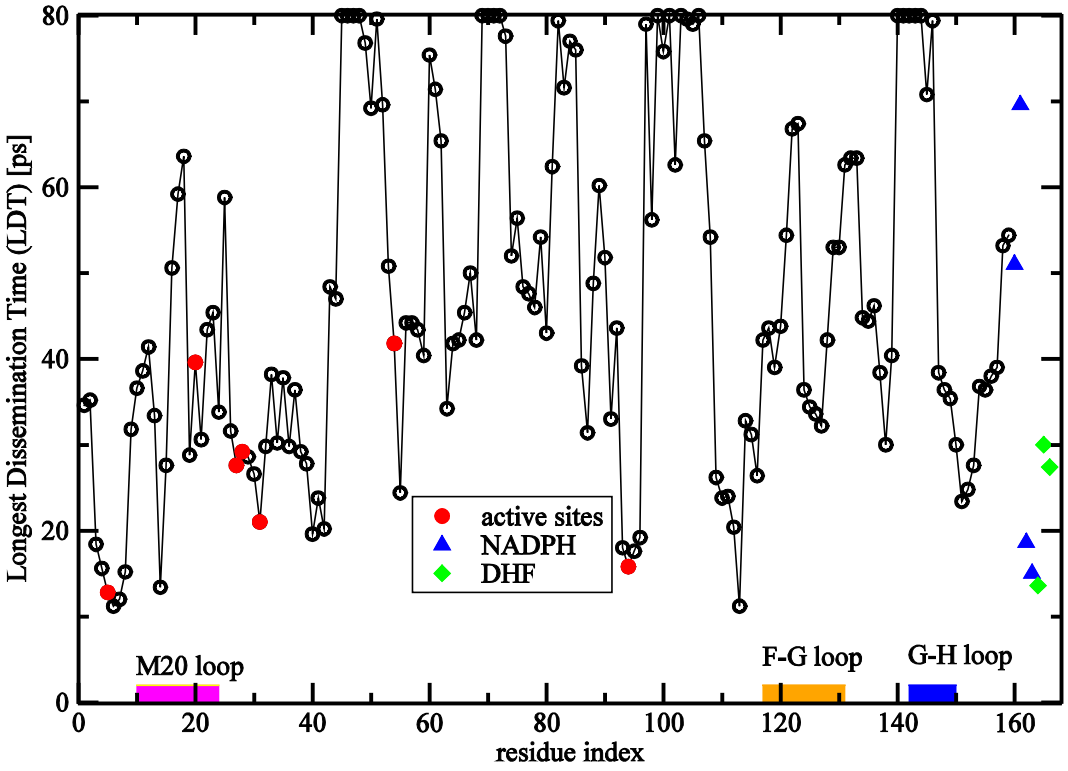

Figure 3 The longest dissemination times (LDTs) (49) of selected indices of sites. An LDT is the slowest response time defined by the longest t_c in the DHFR with 133 evenly distributed impulse forces starting from a C_α atom of a selected residue. Indices 160-163 (in blue triangle) and 164-166 (in green diamond) indicate 4 and 3 groups of NADPH and DHF, respectively (see Figure S2). Index 163 is C4 of nicotinamide group of NADPH and index 164 is C6 of pterin group of DHF; both 163 and 164 are good disseminators with the smallest LDTs among all the perturbation sites (residues). For our studying purpose, we do not concern large LDTs; hence an estimation of LDT larger than 80 ps is shown as 80 ps. The solid red circles indicate the active sites: Ile5, Met20, Asp27, Leu28, Phe31, Leu54 and Ile94 (62) showing relatively small LDT among all.

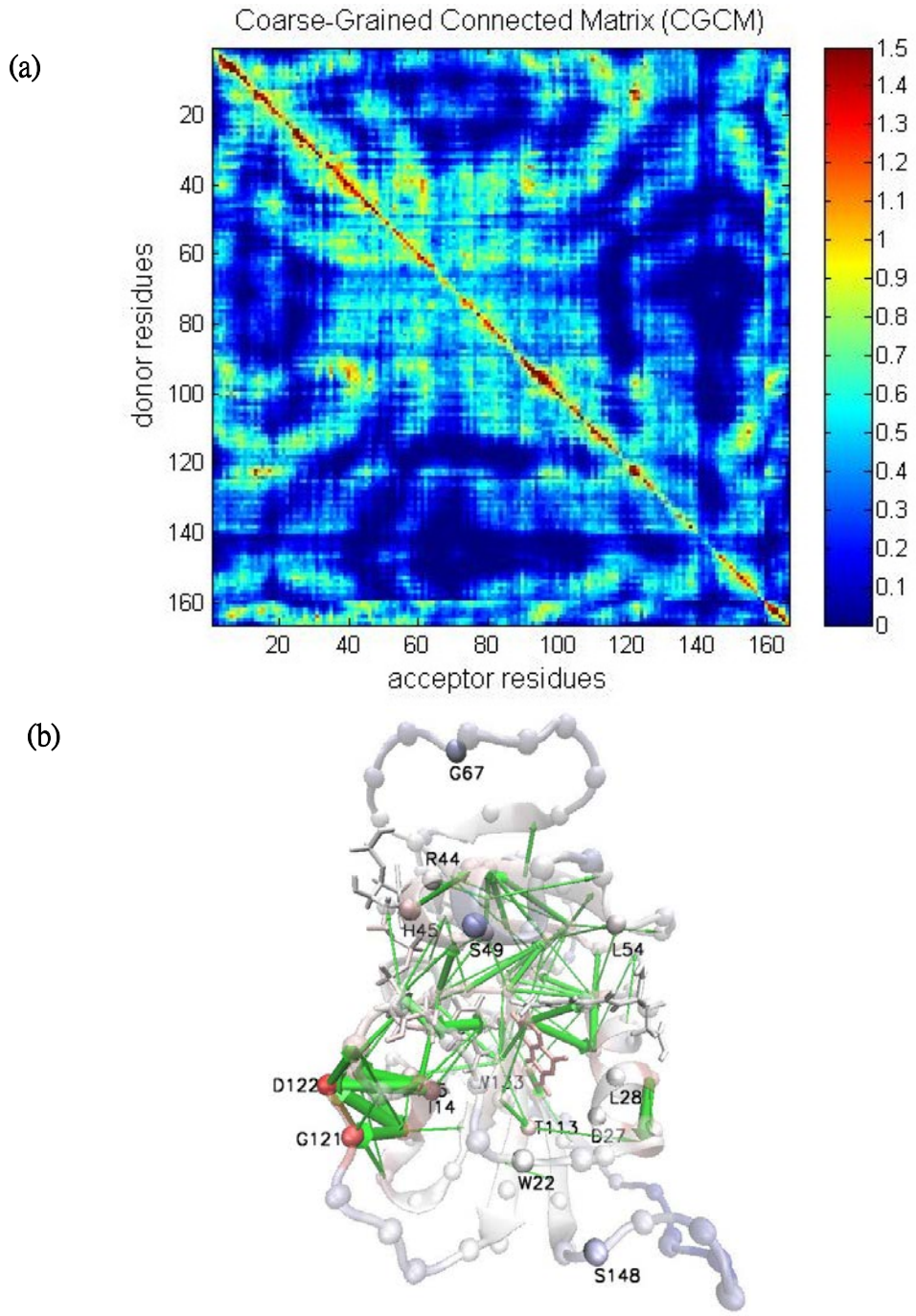

Figure 4 (a) The coarse-grained connection matrix (CGCM) in which each element, $C(R_d, R_a)$, records the counts of mechanical signals propagating from a donor (R_d) residue to an acceptor (R_a) residue, generated by the method described in section 2d. Consequently, a unique "communication score (CS)" is assigned to a selected residue with the highest CGCM off-diagonal element satisfying |index(R_d)-index(R_a)| > 2 pertaining to that residue (either a donor or an acceptor). In other words, the CS of residue *i* is the highest CGCM score of the *i-th* row or the *i-th* column. For example, selecting the residue 122, as a donor, the highest CGCM off-diagonal elements in the 122-th row is $C(122,13) = 1.39$; as an acceptor, the highest CGCM in the 122-th column is $C(13,122) = 1.52$, therefore, the CS of residue 122 is 1.52. **(b)** The CSs color-coded 3D structure of the DHFR is plotted by the software VMD, where cartoon and spheres (C_α atoms) are colored by mapping the minimal CS (dark blue) and the maximal CS to unity (dark red). The green arrows pointing from donors to acceptors

indicate signal propagation pathways, where the thickness represents the size of CS for the pair of interest.

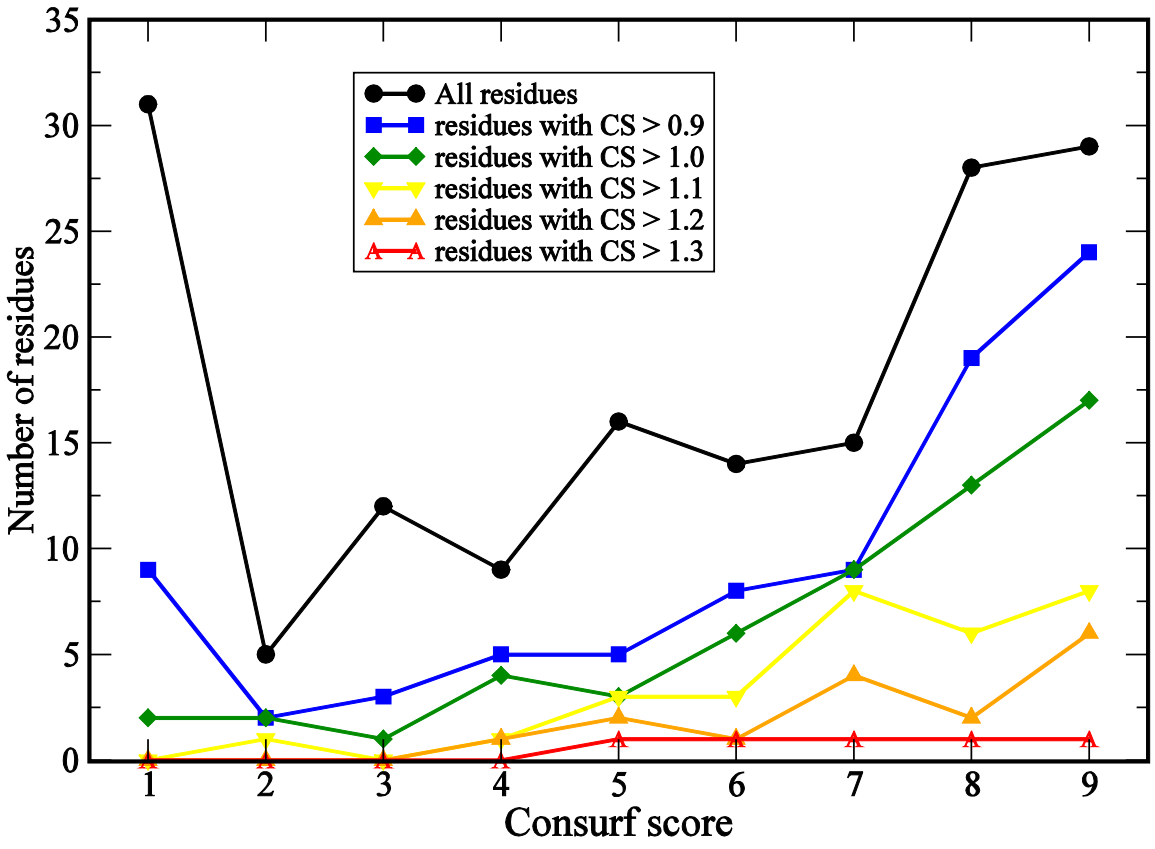

Figure 5 The histograms of conservation scores (Consurf scores; 63) for the residues of different CSs, where score 9 indicates the highest evolutionary conservation and 1 the lowest based on multiple sequence alignment results (63). The black line indicates the distribution for all 159 residues of DHFR. The blue, green, yellow, orange and red lines correspond to the residues with CSs higher than 0.9, 1.0, 1.1, 1.2 and 1.3, respectively.

Table 1. Fold reduction of hydride transfer rates, k_{WT}/k_{mut}, for mutants

WT / mutant	14 I	22 W	27 D	28 L	42 M	44 R	45 H	49 S	54 L	67 G	100 Y	113 T	121 G	122 D	125 F	133 W	148 S
A	40							1.05					5.88[11]	55			1.4
C			79														
D																	0.69
E			3														
F		3		0.24	1.42				47.5[29]							2.6	
G	1036								33[24]		63.3						
H		130															
I									30.7[24]		27.1						
K																	1.36
L						21							1098[10]				
M															42		
N									22.6[24]					23.4			
P													500[10]				
Q							2.8										
S													62.5[11]	37.3			
V	7									1.2		5.8	166[11]				
W					41												
Y				8.7													
-⟨ΔΔG⟩	2.52	1.79	1.64	0.22	1.22	1.83	0.62	0.03	2.08	0.11	2.23	1.05	3.64	2.16	2.24	0.57	0.06
CS	1.22	0.94	0.94	0.83	1.07	1.00	1.12	0.66	1.01	0.69	1.09	1.08	1.41	1.52	1.08	0.70	0.74
D_{act} (Å)	9.5	10.7	8.9	7.5	9.3	13.6	13.0	11.5	10.1	21.0	11.6	10.4	15.4	15.6	12.7	21.3	16.1
ConSurf score	9	9	9	8	9	9	8	9	9	1	8	9	8	9	9	8	1
References	(30)	(25)	(21)	(29)	(11)	(31)	(31)	(27)	(29) (24)	(11)	(27)	(32)	(11) (10)	(22)	(28)	(28)	(23)

By surveying experimental data from the literature (see the References in the bottom row), the fold reduction (Γ) of hydride transfer rates are indexed by the amino acids in mutants at the row and the mutants' original wild-type (WT) amino acids at the columns, where $\Gamma = k_{WT}/k_{mut}$. For example, the I14A mutation causes 40 fold of hydride transfer rate reduction (the most upper left value in the table). The red and blue highlights indicate active sites and perturbation sites, respectively. $\langle \Delta\Delta G \rangle = \langle \Delta G_{WT} - \Delta G_{mutant} \rangle = \langle -RT ln(\Gamma) \rangle$ is the average of a set of changes of free energy difference in the hydride transfer rates between the wild type and its mutants at a given site (average over the rows). CS is the communication score; $\mathbf{D_{act}}$ is the distance between the middle point of N5 and C6 atoms of DHF (active site) and the C_α atom of a mutated site. It is worth noting that many mutants have structural evidence to prove there is no significant conformational changes involved (8, 20, 28-33), as compared to the WT conformation. The ConSurf score (63) represents the conservation of a residue, where score 9 is the most conserved and 1 is the least.

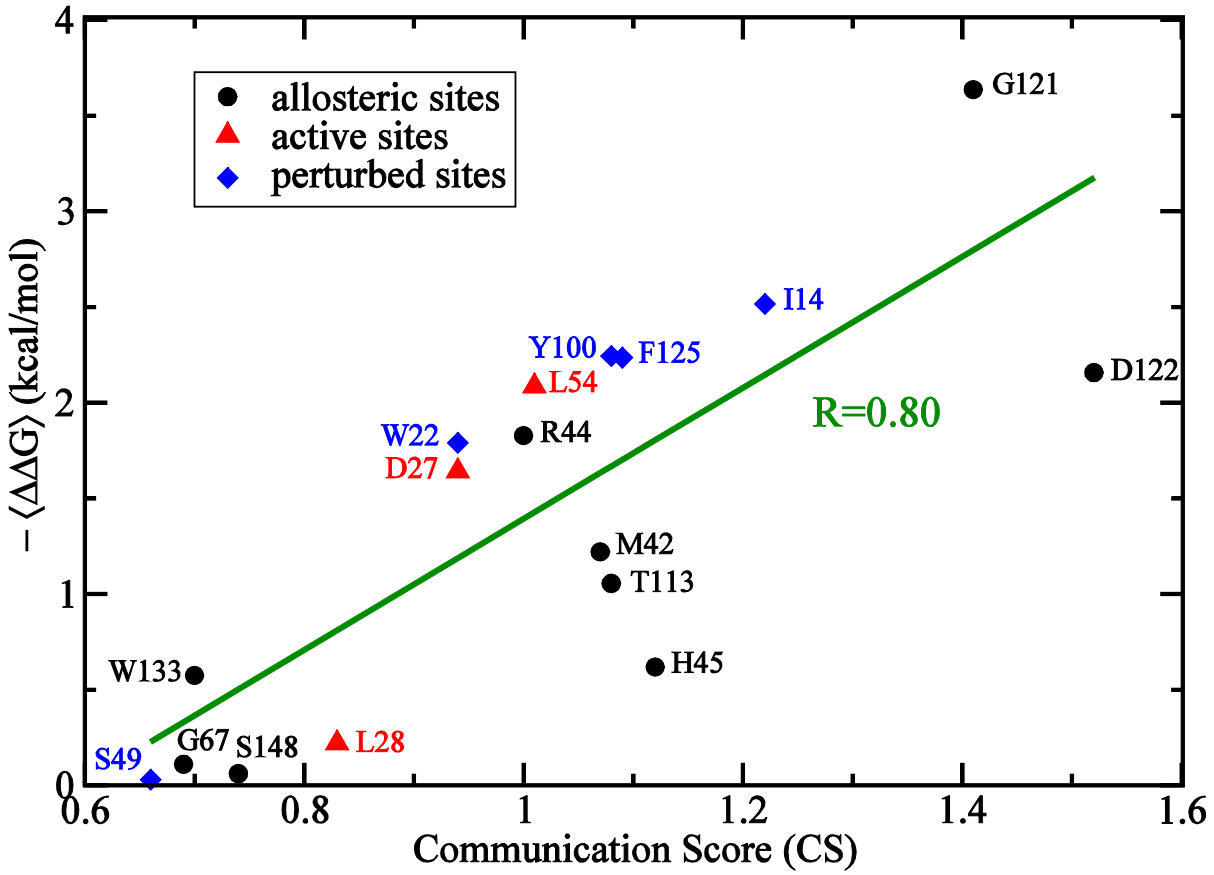

Figure 6 The averaged change in free energy difference of the hydride transfer rates between the wild type and its mutants is plotted against the communication score (CS). The averaged $\langle \Delta\Delta \mathbf{G} \rangle = \langle \Delta \boldsymbol{G_{WT}} - \Delta \boldsymbol{G_{mutant}} \rangle = \langle -RTln(k_{WT}/k_{mut}) \rangle$, where k_{WT} and k_{mut} are the hydride transfer rates of wild type DHFR and its mutant, respectively. $\langle ... \rangle$ denotes the average over different mutants at the same site; the data are listed in **Table 1**. The CS for each residue is determined following the method in section 2d, and CSs for all the residues in DHFR are listed in **Table S1**. The red triangles are active sites, blue diamonds are perturbed sites and black circles are allosteric sites (>15Å from the catalytic center). The dark green line shows linear regression of the data for 17 sites with a correlation coefficient of 0.80.

REFERENCES

1. Benkovic, S. J., and S. Hammes-Schiffer. 2003. A perspective on enzyme catalysis. Science 301:1196-1202.
2. Goodey, N. M., and S. J. Benkovic. 2008. Allosteric regulation and catalysis emerge via a common route. Nature chemical biology 4:474-482.
3. Okazaki, K., and S. Takada. 2008. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. Proceedings of the National Academy of Sciences of the United States of America 105:11182-11187.
4. Monod, J., J. Wyman, and J. P. Changeux. 1965. On the Nature of Allosteric Transitions: A Plausible Model. J Mol Biol 12:88-118.
5. Koshland, D. E., Jr., G. Nemethy, and D. Filmer. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5:365-385.
6. Popovych, N., S. Sun, R. H. Ebright, and C. G. Kalodimos. 2006. Dynamically driven protein allostery. Nature structural & molecular biology 13:831-838.
7. Fierke, C. A., K. A. Johnson, and S. J. Benkovic. 1987. Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli. Biochemistry 26:4085-4092.
8. Boehr, D. D., D. McElheny, H. J. Dyson, and P. E. Wright. 2006. The dynamic energy landscape of dihydrofolate reductase catalysis. Science 313:1638-1642.
9. Yokota, A., H. Takahashi, T. Takenawa, and M. Arai. 2010. Probing the roles of conserved arginine-44 of Escherichia coli dihydrofolate reductase in its function and stability by systematic sequence perturbation analysis. Biochemical and biophysical research communications 391:1703-1707.
10. Cameron, C. E., and S. J. Benkovic. 1997. Evidence for a functional role of the dynamics of glycine-121 of Escherichia coli dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant. Biochemistry 36:15792-15800.
11. Rajagopalan, P. T., S. Lutz, and S. J. Benkovic. 2002. Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: mutational effects on hydride transfer rates. Biochemistry 41:12618-12628.
12. Watney, J. B., P. K. Agarwal, and S. Hammes-Schiffer. 2003. Effect of mutation on enzyme motion in dihydrofolate reductase. Journal of the American Chemical Society 125:3745-3750.
13. Takahashi, H., A. Yokota, T. Takenawa, and M. Iwakura. 2009. Sequence perturbation analysis: addressing amino acid indices to elucidate the C-terminal role of Escherichia coli dihydrofolate reductase. Journal of biochemistry 145:751-762.
14. Garvey, E. P., and C. R. Matthews. 1989. Effects of multiple replacements at a single position on the folding and stability of dihydrofolate reductase from Escherichia coli. Biochemistry 28:2083-2093.
15. O'Neill, J. C., Jr., and C. Robert Matthews. 2000. Localized, stereochemically sensitive

hydrophobic packing in an early folding intermediate of dihydrofolate reductase from Escherichia coli. J Mol Biol 295:737-744.

16. Arai, M., M. Kataoka, K. Kuwajima, C. R. Matthews, and M. Iwakura. 2003. Effects of the Difference in the Unfolded-state Ensemble on the Folding of Escherichia coli Dihydrofolate Reductase. Journal of Molecular Biology 329:779-791.
17. Bystroff, C., S. J. Oatley, and J. Kraut. 1990. Crystal structures of Escherichia coli dihydrofolate reductase: the NADP+ holoenzyme and the folate .cntdot. NADP+ ternary complex. substrate binding and a model for the transition state. Biochemistry 29:3263-3277.
18. Sawaya, M. R., and J. Kraut. 1997. Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. Biochemistry 36:586-603.
19. Radkiewicz, J. L., and C. L. Brooks. 2000. Protein Dynamics in Enzymatic Catalysis: Exploration of Dihydrofolate Reductase. Journal of the American Chemical Society 122:225-231.
20. Osborne, M. J., J. Schnell, S. J. Benkovic, H. J. Dyson, and P. E. Wright. 2001. Backbone Dynamics in Dihydrofolate Reductase Complexes: Role of Loop Flexibility in the Catalytic Mechanism. Biochemistry 40:9846-9859.
21. David, C. L., E. E. Howell, M. F. Farnum, J. E. Villafranca, S. J. Oatley, and J. Kraut. 1992. Structure and function of alternative proton-relay mutants of dihydrofolate reductase. Biochemistry 31:9813-9822.
22. Miller, G. P., and S. J. Benkovic. 1998. Strength of an interloop hydrogen bond determines the kinetic pathway in catalysis by Escherichia coli dihydrofolate reductase. Biochemistry 37:6336-6342.
23. Miller, G. P., D. C. Wahnon, and S. J. Benkovic. 2001. Interloop Contacts Modulate Ligand Cycling during Catalysis byEscherichia coliDihydrofolate Reductase†. Biochemistry 40:867-875.
24. Murphy, D. J., and S. J. Benkovic. 1989. Hydrophobic interactions via mutants of Escherichia coli dihydrofolate reductase: separation of binding and catalysis. Biochemistry 28:3025-3031.
25. Warren, M. S., K. A. Brown, M. F. Farnum, E. E. Howell, and J. Kraut. 1991. Investigation of the functional role of tryptophan-22 in Escherichia coli dihydrofolate reductase by site-directed mutagenesis. Biochemistry 30:11092-11103.
26. Rod, T. H., J. L. Radkiewicz, and C. L. Brooks, 3rd. 2003. Correlated motion and the effect of distal mutations in dihydrofolate reductase. Proceedings of the National Academy of Sciences of the United States of America 100:6980-6985.
27. Adams, J. A., C. A. Fierke, and S. J. Benkovic. 1991. The function of amino acid residues contacting the nicotinamide ring of NADPH in dihydrofolate reductase from Escherichia coli. Biochemistry 30:11046-11054.
28. Singh, P., A. Sen, K. Francis, and A. Kohen. 2014. Extension and limits of the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. Journal of the

American Chemical Society 136:2575-2582.

29. Huang, Z., C. R. Wagner, and S. J. Benkovic. 1994. Nonadditivity of mutational effects at the folate binding site of Escherichia coli dihydrofolate reductase. Biochemistry 33:11576-11585.
30. Stojkovic, V., L. L. Perissinotti, D. Willmer, S. J. Benkovic, and A. Kohen. 2012. Effects of the donor-acceptor distance and dynamics on hydride tunneling in the dihydrofolate reductase catalyzed reaction. Journal of the American Chemical Society 134:1738-1745.
31. Adams, J., K. Johnson, R. Matthews, and S. J. Benkovic. 1989. Effects of distal point-site mutations on the binding and catalysis of dihydrofolate reductase from Escherichia coli. Biochemistry 28:6611-6618.
32. Fierke, C. A., and S. J. Benkovic. 1989. Probing the functional role of threonine-113 of Escherichia coli dihydrofolate reductase for its effect on turnover efficiency, catalysis, and binding. Biochemistry 28:478-486.
33. Boehr, D. D., J. R. Schnell, D. McElheny, S. H. Bae, B. M. Duggan, S. J. Benkovic, H. J. Dyson, and P. E. Wright. 2013. A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. Biochemistry 52:4605-4619.
34. Agarwal, P. K., S. R. Billeter, P. T. Rajagopalan, S. J. Benkovic, and S. Hammes-Schiffer. 2002. Network of coupled promoting motions in enzyme catalysis. Proceedings of the National Academy of Sciences of the United States of America 99:2794-2799.
35. Austin, R., A. Xie, L. van der Meer, B. Redlich, P.-A. Lindgård, H. Frauenfelder, and D. Fu. 2005. Picosecond Thermometer in the Amide I Band of Myoglobin. Physical Review Letters 94.
36. Sato, A., Y. Gao, T. Kitagawa, and Y. Mizutani. 2007. Primary protein response after ligand photodissociation in carbonmonoxy myoglobin. Proceedings of the National Academy of Sciences of the United States of America 104:9627-9632.
37. He, Y., J. Y. Chen, J. R. Knab, W. Zheng, and A. G. Markelz. 2011. Evidence of protein collective motions on the picosecond timescale. Biophys J 100:1058-1065.
38. Acbas, G., K. A. Niessen, E. H. Snell, and A. G. Markelz. 2014. Optical measurements of long-range protein vibrations. Nature communications 5:3076.
39. Gnanasekaran, R., J. K. Agbo, and D. M. Leitner. 2011. Communication maps computed for homodimeric hemoglobin: computational study of water-mediated energy transport in proteins. The Journal of chemical physics 135:065103.
40. Leitner, D. M. 2009. Frequency-resolved communication maps for proteins and other nanoscale materials. The Journal of chemical physics 130:195101.
41. Ota, N., and D. A. Agard. 2005. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. J Mol Biol 351:345-354.
42. Chennubhotla, C., and I. Bahar. 2006. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. Molecular systems biology 2:36.
43. Ghosh, A., and S. Vishveshwara. 2007. A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis.

Proceedings of the National Academy of Sciences of the United States of America 104:15711-15716.

44. Kong, Y., and M. Karplus. 2009. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 74:145-154.

45. Sharp, K., and J. J. Skinner. 2006. Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. Proteins 65:347-361.

46. Lockless, S. W., and R. Ranganathan. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286:295-299.

47. Chennubhotla, C., and I. Bahar. 2007. Signal propagation in proteins and relation to equilibrium fluctuations. PLoS computational biology 3:1716-1726.

48. Yu, X., and D. M. Leitner. 2005. Heat flow in proteins: computation of thermal transport coefficients. The Journal of chemical physics 122:54902.

49. Yang, L. W., A. Kitao, B. C. Huang, and N. Go. 2014. Ligand-induced protein responses and mechanical signal propagation described by linear response theories. Biophys J 107:1415-1425.

50. Brooks, B., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proceedings of the National Academy of Sciences of the United States of America 80:6571-6575.

51. Go, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proceedings of the National Academy of Sciences of the United States of America 80:3696-3700.

52. Chandrasekhar, S. 1943. Stochastic Problems in Physics and Astronomy. Reviews of Modern Physics 15:1-89.

53. Schotte, F., M. H. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, and P. A. Anfinrud. 2003. Watching a protein as it functions with 150-ps time-resolved X-ray crystallography. Science 300:1944-1947.

56. Huang, X., and S. G. Boxer. 1994. Discovery of new ligand binding pathways in myoglobin by random mutagenesis. Nature structural biology 1:226-229.

57. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. Journal of computational chemistry 26:1781-1802.

58. Best, R. B., X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. Journal of chemical theory and computation 8:3257-3273.

59. Vanommeslaeghe, K., and A. D. MacKerell, Jr. 2012. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. Journal of chemical information and modeling 52:3144-3154.

60. Vanommeslaeghe, K., E. P. Raman, and A. D. MacKerell, Jr. 2012. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. Journal of chemical information and modeling 52:3155-3168.

61. Hayward, S., A. Kitao, F. Hirata, and N. Go. 1993. Effect of solvent on collective motions in globular protein. J Mol Biol 234:1207-1217.

62. Brown, K. A., and J. Kraut. 1992. Exploring the molecular mechanism of dihydrofolate reductase. Faraday Discussions 93:217-.

63. Ashkenazy, H., E. Erez, E. Martz, T. Pupko, and N. Ben-Tal. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic acids research 38:W529-533.

64. Cooper A, Dryden DT (1984) Coupling of global and local vibrational modes in dynamic allostery of proteins. Eur Biophys J. 11:103-109.

65. Faruck Morcos, Andrea Pagnani, Bryan Lunt, Arianna Bertolino, Debora S. Marks, Chris Sander, Riccardo Zecchina, José N. Onuchic, Terence Hwa, and Martin Weigt (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. 108: E1293-E1301

66. Klinman, J. P., and A. Kohen. 2014. Evolutionary Aspects of Enzyme Dynamics. The Journal of biological chemistry 289:30205-30212.

67. Li,H., Chang,YY, Lee,JY, Bahar,I. and Yang,LW. (2017) DynOmics: dynamics of structural. proteome and beyond. Nucleic Acids Res., 45, W374-W380.

68. Yang, L.W. and Bahar, I. (2005) Coupling between catalytic site and collective dynamics: a. requirement for mechanochemical activity of enzymes. Structure, 13, 893-904.

69. Yang, L.W., Eyal, E., Bahar, I. and Kitao, A. (2009) Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. Bioinformatics, 25, 606-614.

70. Hongchun Li, Shun Sakuraba, Aravind Chandrasekaran, and Lee-Wei Yang. Molecular Binding. Sites Are Located Near the Interface of Intrinsic Dynamics Domains (IDDs). J. Chem. Inf. Model. 2014, 54(8), 2275-2285

71. Chandrasekaran, Aravind, Chan, Justin, Lim, Carmay, Yang, Lee-Wei (2016) Protein dynamics. and contact topology reveal protein-DNA binding orientation" Journal of Chemical Theory and Computation

72. Ivet Bahar, Ali Rana Atilgan, Melik C. Demirel, and Burak Erman (1998) Vibrational Dynamics. of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. Phys. Rev. Lett. 80, 2733

SUPPORTING INFORMATION

Movie of ICCs of DHFR

The active sites and perturbed sites

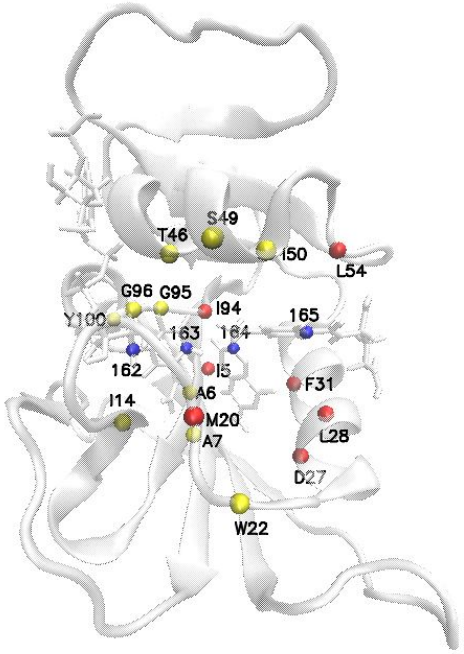

Figure S1. The 21 spheres are perturbed by impulse forces, where the red spheres are active sites: ILE5, MET20, ASP27, LEU28, PHE31, LEU54 and ILE94; yellow spheres are residues within 7 **Å** from the catalytic center: ALA6, ALA7, ILE14, TRP22, THR46, SER49, ILE50, GLY95, GLY96, TYR100. The four blue spheres are the perturbed sites at cofactor and substrate, which is the first Carbon (site 162) of the NADPH at the ribose close to the nicotinamide, the C4 (site 163) of the NADPH at the nicotinamide close to the donor hydride, the C6 (site 164) of the pterin group of the DHF and C15 (site 165) of the benzoyl group of the DHF.

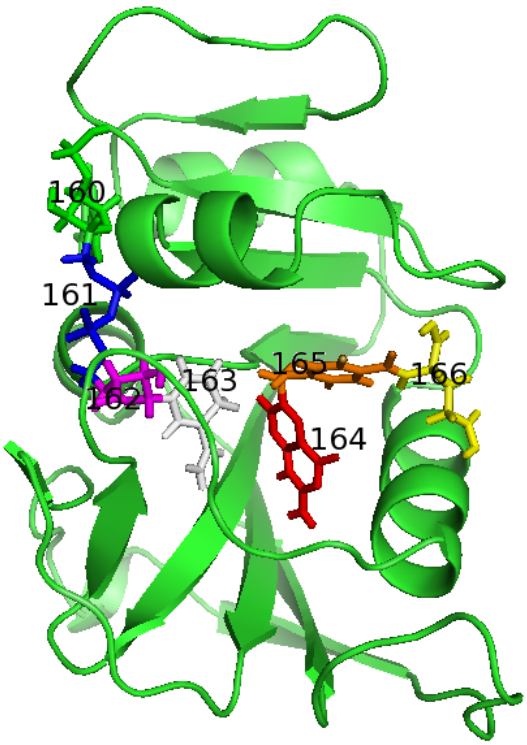

Figure S2. The coarse-grained NADPH and dihydrofolate. The NADPH is divided into four groups, 160-163, colored by green, blue, purple and white, respectively. The dihydrofolate is divided into three groups 164-166 colored by red, orange and yellow.

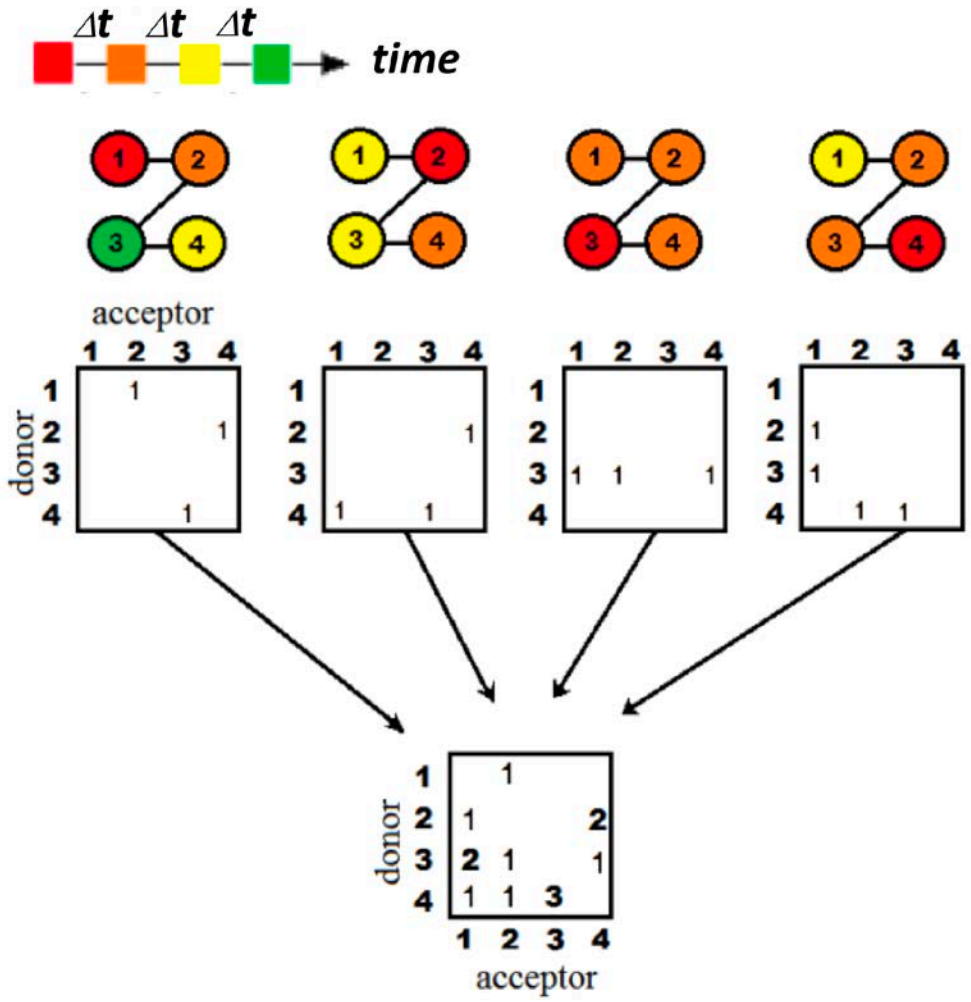

Figure S3 Recording the signals into the connection matrices. The colors indicate a time sequence of characteristic times. The circles with number linked by solid line are the model protein sequences, where the number is the atomic index. Given different perturbed forces, $\vec{k}_j$, the signals propagation pathways were varied. The squares are the matrices used to record the signals. A connected pair satisfying criteria was recorded into an element of a connection matrix by adding a count. Following eq. (2), after summation over all matrices with different $\vec{k}_j$, we have the full-atom connected matrix $F(a_d, a_r)$.

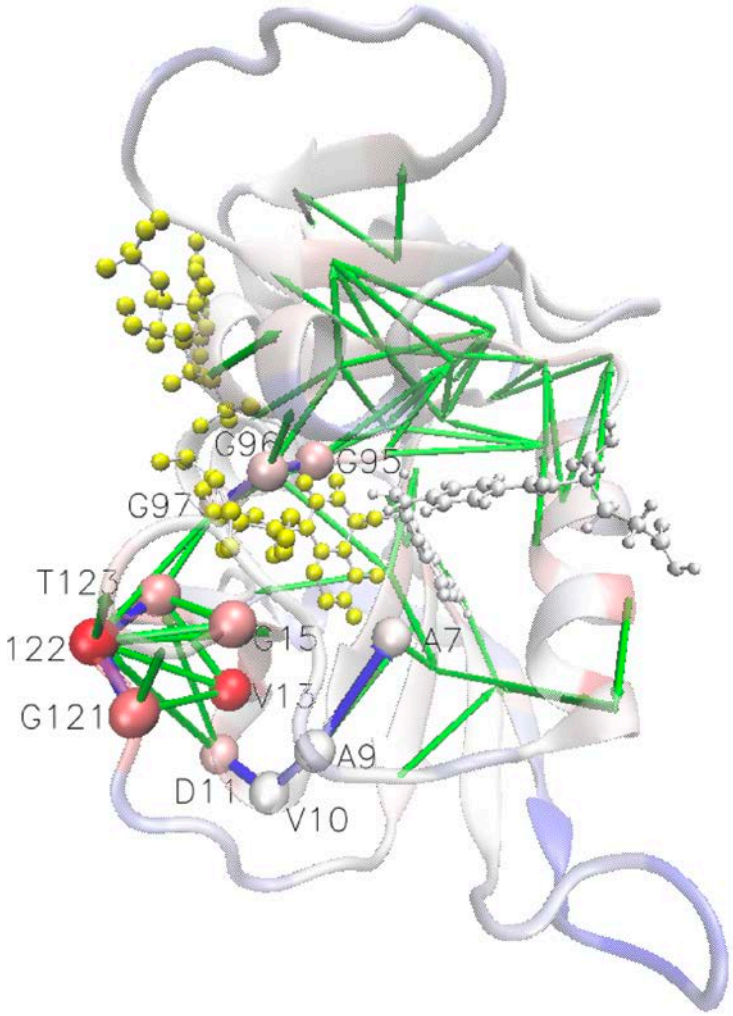

Figure S4 The communication pathways from a remote allosteric site G121 to sites close to catalytic centers. The pathway through the upper-middle (in blue) G121(0.88) →D122(1.00) →T123(0.81) → G97(0.70) → G96(0.73) →G95(0.76) ; The pathway through underneath the cofactor (in blue) : G121(0.88) → D122(1.00) →D11(0.75) → V10 (0.60) →A9(0.63) →A7(0.67). The values in the parentheses are the communication scores (CSs); the direction of signal propagation from one residue (donor) to the other (acceptor) in a residue pair can be assigned is because the communication is asymmetric in CGCM, where $C(R_i, R_j) \neq C(R_j, R_i)$ (although the two numbers are usually close). The direction of signal is said from residue i to j if $C(R_i, R_j) > C(R_j, R_i)$.

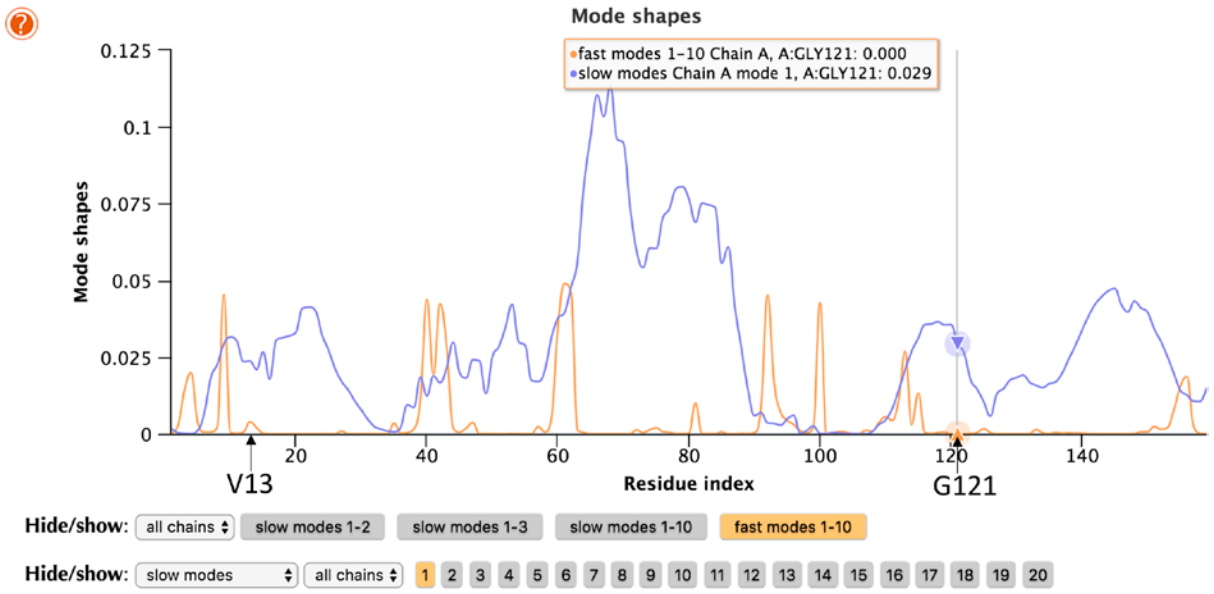

Figure S5 The ENM slowest mode (orange) and the average of the fastest 10 ENM mode (blue) for ternary DHFR complex (1RX6) are plotted against residue index. Frequent communicators V13-G121 are marked by black arrows. The analysis and resulting figure are taken from the DynOmics website (Li et al., 2017; https://dyn.life.nthu.edu.tw/oENM/).

Table S1 The communication scores (CSs) of residues in the DHFR

rid / score		rid / score		rid / score		rid / score		rid / score		rid / score		rid / score		rid / score	
1	0.72	**21**	0.73	**41**	1.14	**61**	1.18	**81**	0.88	**101**	0.96	**121**	1.41	**141**	0.39
2	0.76	**22**	0.94	**42**	1.07	**62**	1.03	**82**	0.92	**102**	0.88	**122**	1.52	**142**	0.33
3	1.04	**23**	0.81	**43**	1.02	**63**	0.96	**83**	0.84	**103**	0.85	**123**	1.29	**143**	0.49
4	1.22	**24**	0.98	**44**	1	**64**	0.88	**84**	0.81	**104**	0.79	**124**	0.86	**144**	0.33
5	1.03	**25**	1.06	**45**	1.12	**65**	0.66	**85**	0.83	**105**	0.88	**125**	1.08	**145**	0.58
6	1.1	**26**	1.31	**46**	1.14	**66**	0.59	**86**	0.76	**106**	0.92	**126**	0.93	**146**	0.54
7	1.08	**27**	0.94	**47**	1.06	**67**	0.69	**87**	0.66	**107**	1.02	**127**	0.81	**147**	0.74
8	1.03	**28**	0.83	**48**	1.12	**68**	0.68	**88**	0.72	**108**	0.7	**128**	0.64	**148**	0.74
9	0.98	**29**	1.31	**49**	0.66	**69**	0.72	**89**	0.59	**109**	0.97	**129**	0.67	**149**	0.7
10	0.93	**30**	0.85	**50**	0.73	**70**	0.69	**90**	0.94	**110**	1.22	**130**	0.54	**150**	0.87
11	1.12	**31**	1.22	**51**	0.85	**71**	0.69	**91**	1.01	**111**	1.23	**131**	0.63	**151**	0.89
12	0.91	**32**	1	**52**	0.61	**72**	0.83	**92**	1.17	**112**	0.94	**132**	0.63	**152**	0.98
13	1.52	**33**	0.78	**53**	1.06	**73**	1.07	**93**	1.11	**113**	1.08	**133**	0.7	**153**	0.94
14	1.22	**34**	1.12	**54**	1.01	**74**	0.87	**94**	1.18	**114**	0.98	**134**	0.72	**154**	0.85
15	1.29	**35**	1.21	**55**	0.81	**75**	0.95	**95**	1.21	**115**	0.98	**135**	0.84	**155**	1.23
16	1.01	**36**	0.79	**56**	1.03	**76**	0.7	**96**	1.14	**116**	0.9	**136**	0.61	**156**	0.9
17	1.17	**37**	1.05	**57**	1.03	**77**	0.84	**97**	1.07	**117**	0.9	**137**	0.68	**157**	0.85
18	0.92	**38**	1.12	**58**	1.08	**78**	0.82	**98**	0.97	**118**	0.71	**138**	1.02	**158**	0.64
19	0.97	**39**	1.18	**59**	0.89	**79**	0.92	**99**	1.02	**119**	0.78	**139**	0.71	**159**	0.72
20	0.97	**40**	1.21	**60**	1.07	**80**	0.95	**100**	1.09	**120**	0.67	**140**	0.65		

The bold texts indicate the residue indices (rid).

Table S2 The ConSurf scores of residues in DHFR

ind / score		ind / score		ind/ score		ind / score		ind / score		ind / score		ind / score		ind / score	
1	7	**21**	8	**41**	6	**61**	8	**81**	8	**101**	1	**121**	8	**141**	7
2	7	**22**	9	**42**	9	**62**	4	**82**	3	**102**	5	**122**	9	**142**	1
3	8	**23**	6	**43**	9	**63**	8	**83**	1	**103**	3	**123**	8	**143**	5
4	4	**24**	9	**44**	9	**64**	8	**84**	1	**104**	6	**124**	1	**144**	8
5	8	**25**	9	**45**	8	**65**	3	**85**	4	**105**	6	**125**	9	**145**	5
6	7	**26**	6	**46**	9	**66**	1	**86**	1	**106**	1	**126**	9	**146**	3
7	9	**27**	9	**47**	1	**67**	1	**87**	1	**107**	8	**127**	3	**147**	8
8	6	**28**	8	**48**	7	**68**	1	**88**	1	**108**	5	**128**	1	**148**	1
9	8	**29**	5	**49**	9	**69**	4	**89**	1	**109**	5	**129**	2	**149**	5
10	1	**30**	6	**50**	8	**70**	1	**90**	8	**110**	7	**130**	1	**150**	1
11	8	**31**	9	**51**	9	**71**	6	**91**	4	**111**	7	**131**	1	**151**	5
12	5	**32**	9	**52**	7	**72**	5	**92**	7	**112**	6	**132**	1	**152**	3
13	7	**33**	3	**53**	9	**73**	1	**93**	6	**113**	9	**133**	8	**153**	8
14	9	**34**	2	**54**	9	**74**	6	**94**	8	**114**	1	**134**	1	**154**	1
15	9	**35**	9	**55**	9	**75**	6	**95**	9	**115**	8	**135**	3	**155**	5
16	4	**36**	5	**56**	6	**76**	3	**96**	9	**116**	4	**136**	3	**156**	5
17	5	**37**	8	**57**	9	**77**	8	**97**	8	**117**	4	**137**	5	**157**	2
18	9	**38**	7	**58**	3	**78**	2	**98**	4	**118**	1	**138**	2	**158**	9
19	1	**39**	7	**59**	9	**79**	1	**99**	6	**119**	5	**139**	7	**159**	7
20	8	**40**	7	**60**	8	**80**	1	**100**	8	**120**	3	**140**	1		

The bold texts indicate the residue indices (ind).

Table S3. Contribution per perturbation site to C(13,121) and C(13,122)

perturbed sites	C(13,121)	C(13,122)
5	0.0788	0.1015
6	0.0606	0.0415
7	0.0673	0.0757
14	0.0428	0.0265
20	0.0046	0.0114
22	0.0271	0.0134
27	0.0646	0.0836
28	0.1159	0.1455
31	0.0508	0.0252
46	0.0904	0.0837
49	0.0195	0.039
50	0.012	0.0349
54	0.0816	0.0905
94	0.0458	0.1476
95	0.0713	0.0745
96	0.0315	0.0401
100	0.1662	0.1053
162	0.1086	0.0932
163	0.1028	0.1435
164	0.1687	0.1328
165	0.0073	0.0107
Total score	1.4184	1.5201

SUPPORTING METHODS

Time-independent linear response theory

As described in our previous work (Yang et al., 2014; Ref 49), the time trajectory of changes in atomic position *i* can be described by the interplay between a velocity-position time correlation function and time-dependent perturbation forces summed over the atoms that are coupled with the atom *i*[1].

$$\langle \Delta\vec{\mathbf{r}}_{\mathrm{i}}(\mathrm{t})\rangle_{\mathrm{f}} = \frac{1}{k_B T}\int_0^t dt' \sum \langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t)\rangle_0 f_j(t-t') \qquad \text{(S1)}$$

where $\Delta\vec{\mathbf{r}}_{\mathrm{j}}$ is the deviation from the mean of atom *j* due to the external force $\vec{\mathbf{f}}_j$; $\Delta\dot{\vec{r}}_j(0)$ is the velocity of atom *j*; k_B and *T* are the Boltzmann constant and temperature, respectively, $\langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t)\rangle_0$ is the velocity-position time-correlation function sampled in the absence of perturbations (noted by subscript "0"), which can be expressed in the normal-mode space, where modes are treated as independent 1-D harmonic oscillators under solvent damping using the Langevin equation[2]. The detail derivation is referred to our previous work[1].

The time-dependent linear response theory using impulse forces (IF-tdLRT)

Substitute the time-dependent force $f_j(t-t')$ in eq. (S1) with an impulse force, presented by a delta function $\vec{k}_j$, the format for the IF-tdLRT is

$$\langle \Delta\vec{r}_{\mathrm{i}}(\mathrm{t})\rangle_{\mathrm{f}} = \frac{1}{k_B T}\sum \langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t)\rangle_0 \vec{k}_j \qquad \text{(S2)}$$

Per Chandrasekhar's description of damped harmonic oscillators[2], when $2\omega_m < \beta$, it can be derived that

$$\langle \Delta\vec{r}_i(t)\rangle_f = \frac{1}{k_B T}\sum_j \left(\sum_m \alpha_{jm}\alpha_{im}\frac{2\langle\sigma_m(0)^2\rangle\omega_m{}^2}{\beta_1} e^{-\beta t/2}\sinh\left(\frac{\beta_1 t}{2}\right)\right)\vec{k}_j \qquad \text{(S3)}$$

when $2\omega_m > \beta$ (let $\beta_1' = \beta_1 i$; note that $\sinh\theta = -i\sin i\theta$), it is

$$\langle \Delta\vec{r}_i(t)\rangle_f = \frac{1}{k_B T}\sum_j \left(\sum_m \alpha_{jm}\alpha_{im}\frac{2\langle\sigma_m(0)^2\rangle\omega_m{}^2}{\beta_1{}'} e^{-\beta t/2}\sin\left(\frac{\beta_1{}' t}{2}\right)\right)\vec{k}_j \qquad \text{(S4)}$$

SUPPORTING REFERENCE

1. L. W. Yang, A. Kitao, B. C. Huang and N. Go, *Biophys J*, 2014, 107, 1415-1425.

2. S. Chandrasekhar, *Reviews of Modern Physics*, 1943, 15, 1-89.