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*

¶ Authors who have equal contributions

* To whom correspondence should be addressed: Lee-Wei Yang. 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 to remote parts in a molecule, and hence a rising interest to identify communication pathways within such a network. Recently, we have developed the time-dependent linear response theories (td-LRTs), which shown to well interpret the ligand photodissociation relaxation dynamics of myoglobin. In this article, we applied the td-LRT with impulse force to investigate the allostery on an enzyme system - the dihydrofolate reductase (DHFR), which catalyze the hydride transfer (HT) reaction - the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using the cofactor nicotinamide adenine dinucleotide phosphate (NADPH). We developed a connection matrix to record the counts of mechanical signals propagating through donor-acceptor pairs launched from the active sites with the td-LRT. Through the connection matrix, the communication score (CS) of each residue is assigned, which quantify how often the signals propagate through a residue in its most accessible way. The sites with high CS are termed "intramolecular communication centers (ICCs)", which were shown to be highly conservative. Furthermore, we showed that the HT rate reduction by a single residue mutation is highly correlated with its CS. Especially, the mysterious 200-fold HT rate reduction of the distal mutation G121V from kinetic isotope effect experiment is well described by our model with its ultra-high CS which means the G121 participating on the signal propagation process most frequently. On the other hand, several mutants: G67V, S148A and W133F causing minor HT rate reduction were described by its low CS, which means fewer signals through those sites. We proposed that the fold of HT rate reduction for a given mutant is an exponential function of its CS, which suggests that the dynamics signal propagation in the wild type play a background role to facilitate the HT quantitatively.

INTRODUCTION

The protein structures, dynamics and functions have been intensively studied for several decades. One of the most challenging parts of the task is to estimate the synergistic effects of mutations far from the binding pockets on the enzyme activity – known as the effect of allostery (1-3). Traditionally, allostery is considered to be driven by noticeable conformational changes. Such like, the pre-existing model (the MWC model) (4) and the KNF model (the induced fit model) (5) were used to explain the allosteric regulation of ligand binding. However, recently it is suggested that the allostery may not involve significant conformational changes. One such example is the cyclic adenosine monophosphate (cAMP) binding to the dimeric catabolite activator protein (CAP), a transcriptional activator (6). Kalodimos's group used the CPMG technique to measure the conformational exchange between states, and found that the binding of first ligand actually affects the microsecond-millisecond dynamics of the unbound subunit, which cause the negative cooperativity of the second cAMP binding (6). It is concluded that the allostery can be driven by protein dynamics without the need of notable conformational changes.

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, 19 Å away from the catalytic center, can reduce the hydride transfer rate by 200 folds (10, 11, 26) while the mutant G67V (11), S148A (7, 23) and W133F (28) impact the activity only minimally. The non-additive property of hydride transfer rate reduction of double mutant data involving positions M42 and G121 further suggest that a coupled dynamics motion of these distal residues enhances crossing of the chemical reaction barrier (11, 26). The QM/MM study also suggests that when a reaction progresses from the reactants to products through chemicals' and protein's transition states; a network of "coupled promoting motions" could facilitate the chemistry in DHFR (12, 34). Furthermore, a recent NMR experiment shows that dynamics on the picosecond-nanosecond and microsecond-millisecond time scales (Using CPMG RD technique) are found significantly changed in the G121V mutant of DHFR, owing to the mutant inducing impaired residue network inside DHFR, which decreases dynamical signal propagation (33).

From NMR evidences, those 'invisible' protein motion ranging from *ps-ns* to *s-ms* are suggested to affect the hydride transfer rate (8). Optical experiments also show that low frequency collective vibrational excitations exist in *ps-ns* (35-38). It is still an extensively studied issue how the enzyme exploit the vibrational excitations through the signal propagation to catalyze the chemical reaction. In the past decade, several theoretical methods were proposed to elucidate intramolecular signal propagation pathways (39-48). Generally, the methods can be divided into two categories –structural/dynamical equilibrium (42-44, 46, 47) or non-equilibrium approaches (39-41, 45, 48). We give a short review in this proposal as well as in our previous work (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 serve 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-natural environment causes the results unreliable. 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 about 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 homodimeric 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. Leitner's approach is methodologically closer to ours than others. Notably, Ranganathan group proposed a correlated mutation method called the "statistical coupling analysis (SCA)" to identity the co-evolved residues (46). This approach however requires many homologous sequences to secure the statistical significance of calculations of wild type and perturbed "energies". (However, a statistical coupling site W133 shows no obvious effect in hydride transfer of DHFR (28))

Beyond previous mentioned theories, recently we have developed a novel theory called the time-dependent linear response theories (td-LRT) to investigate the dynamics of protein motions subject to external perturbations (e.g. ligand dissociation) (49). The method combining the normal mode analysis (NMA) (50, 51) and damped harmonic driven Langevin dynamics (52) has been shown to well describe site-specific residue relaxation in carbonmonoxy myoglobin (Mb) characterized by time-resolved UV Resonance Raman (UVRR) spectroscopy (36). Using td-LRT, we are able to monitor a two-staged relaxation where the slower relaxation occurs from 4.4 to 81.2 ps (49) and the time constants characterized for a couple of aromatic residues agree with those observed by UVRR spectrometry and time-resolved X-ray crystallography (36, 53). The faster 'early responses', triggered as early as 400 femtoseconds, can be best described by the theory when impulse forces are used. Also, the sound speed estimated at 4-7 Å/ps in Mb is compared favorably with that in water (14 Å/ps) and concentration-dependent sound velocity in ~0.1 M protein solution (54, 55). Furthermore, we identify several residues in Mb as 'disseminators' that propagate the signals faster than others and are found to be 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). For example, disseminators L29, A110, V68, I107, I30, L72, S108, F106, G73, E109 are those that have the shortest disseminating time within Mb. We redid the disseminator analysis on the point mutation L29S *in-silico* and found only residues V68, I107, L72, S108 and G73 are still

among the top 10 disseminators; while analysis on another point mutation D60G found that a higher number of wild-type disseminators can be retained in the top 10 disseminators (namely V68, L29, A110, I107, L72, G73, F106). The rebinding kinetics experiment showed that D60G impacts the binding behavior minimally while L29S dramatically changed the rebinding curve (49, 56). The quantitative agreement of our methods with several experiments (36, 53, 56) is an important distinction of LRT from other methods.

In this article, we further apply the td-LRT theory with the impulse force to investigate the dynamic driven allosteric regulation. We propose a new method to track the atomic level dynamic signal propagation pathway, and consequently the recorded signals are coarse-grained to extract the frequently through residues termed as the intramolecular communication centers (ICCs). The evolutional conservation and functional important of the ICCs are also examined. Finally, the possible 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.

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 Param-Chem 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 Param Chem. 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, the forces of atoms are extracted. Consequently, the full atom (*N=2615* atoms including all the hydrogen atoms) normal-mode analysis is carried out by diagonalizing the hessian matrix constructed by the second derivative of the potential energy with respect to 3N mass-weighting Cartesian coordinate. Further detail steps of the normal-mode analysis are referred to our previous work [Biophysics J. 107, 1415(2014)] and its supporting material (49).

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

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

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

, 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 $\langle \Delta\dot{\vec{r}}_j(0)\Delta\vec{r}_i(t')\rangle_0$ with the independent harmonic oscillators using the normal mode analysis (50, 51, 61) under solvent damping by solving the Langevin equation (52). The derivation can be found in SI 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.
In this article, we define a discrete time step dt to evaluate eq. (1), where dt is chosen to 0.2 ps such that signals propagating within few Å (interatomic range) can be recorded. Following eq. (1), the *i-th* atom relaxes to its original equilibrium position after a temporal reconfiguration in response to the perturbation; "the characteristic time" of atom *i*, t_c^i, characterizing the signal propagation process is defined at the time when the *i-th* atom reaches the maximum displacement. It was found that t_c^i is independent of the magnitude of $\vec{k}_j$ but varies versus the direction and location of it. For examining the robustness of several measurements through this article, 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°; 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 dissemation 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 through the entire enzyme from a residue; in another word, to find the longest response time for all possible signals starting from the specific residue. Technically, as the method proposed in our previous work (49), the force ensemble defined in section 2b was applying at certain C_α of residue *s* labeled by *j*. Consequently, a set of characteristic times consist of $133 \times N$ $t_c^i(\Omega_s)$ were generated. A subset which atom *i* is belong to *C* atom, 133×159 $t_c^i(\Omega_s)$ are extracted, and the longest one of the subset is termed as the longest dissemination time (LDT) of residue s. The residues with short LDT are called the "disseminators" (49). We will calculate the LDT of each residue in DHFR by perturbing the corresponding C_α atom, also the C6 atom of the NADPH and N5 atom of the DHF. And compare these calculated disseminators with the active site residues reported by reference (62).

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

Although td-LRT provide the characteristic time $t_c^i(\theta_j, \phi_j)$ to characterize the dynamical signal propagation, a method that trace the signal propagation pathway and identify which residue is important on that pathway is still needed. In this section, we introduce "the Coarse-Grain Connection Matrix (CGCM)" to record the signal propagations between residues. Consequently, we extract "the intramolecular communication centers (ICCs)" - the residues on frequently used 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 pair. There are two connection criteria to determining the atom connection between a donor, a_d, and an acceptor, a_r: First, given an impulse force exerted on the j-th C_α atom, $\vec{k}_j$, the atoms with characteristic times that differ by dt, are connected such that, $t_c^{a_r} - t_c^{a_d} = dt$, where the fixed dt the same as the modeling discrete time step, 0.2 ps. Figure S3 demonstrates how the connection signal is recorded into the connection matrix. In order to maintain the causality of signal propagation, the candidates satisfy the first criteria are filtered by the Second criteria in addition- the angle between the vector *j*-th C_α to a_r and *j*-th C_α to a_d should be less than 90 degree. It is possible that a donor atom connects many acceptor atoms, or many donors connect to an acceptor atom. Consequently, in order to quantitatively recognize how frequently a residue participate in the

signal propagation, we define a full-atom connected matrix $F(a_d, a_r)$ account for the number of the connection that the donor, a_d, connects to the acceptor, a_r

$$F(a_d, a_r) = \sum_j \langle \delta(dt - [t_c^{a_r}(\Omega_j) - t_c^{a_d}(\Omega_j)]) \rangle_\Omega, \qquad (2)$$

where j runs over the selected C_α atoms, $\delta(x)$ is a Delta function equal to zero for any nonzero x, and $\delta(0) = 1$. Ω_j are forces orientation angles defined in section 2b. Fig.4 illustrates how to build a full atom connected matrix. With the full-atom F matrix, we further define the residue level CGCM by summing the counts belong to each residue pair, $C(R_d, R_r) = \sum_{a_d \in R_d} \sum_{a_r \in R_r} \frac{F(a_d, a_r)}{N_d N_r}$, where R_d and R_r are donor and acceptor residues, respectively; N_d and N_r are the number of atom in donor and acceptor residue, respectively. To modeling the signals propagation process during the hydride transfer reaction, we perturbed 21 sites including C 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 S2). Consequently, 133 evenly distributed impulse forces defined in section 2b are applying at each perturbed site, all signals were sum up to one CGCM, $C(R_d, R_r)$, normalized by 133 directions. In the CGCM, the diagonal elements, indicating sequentially adjacent residues, are relative high in $C(R_d, R_r)$ scores, which is intuitive but less meaningful in terms of the allostery. For the off-diagonal elements satisfying $|R_d - R_r| > 2$, they provide information on the signal propagation between long range contacts (say, within or between secondary structures). Several hot spots with higher CGCM score indicate the residue pairs on frequently used pathways communicating dynamic signals. Then, a unique "**communication score (CS)**" assigned to each residue is 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 SI table 7.

2e. Examining evolutional significance of the ICCs

The evolutional importance of residues is compared with its CS. The sequence conservation of residues are calculated by multiple sequence alignment conducted by the ConSurf web site (63-65), where residues are categorized into 9 conservation levels by "the ConSurf score" from 1 (most diverse) to 9 (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. Examining the coupling between mechanical/communicational significance of allosteric sites and their impacts on hydride transfer rate of DHFR upon site-directed mutagenesis

The forward hydride transfer rate (for chemical step, k_{hyd}) of wild type DHFR defined as k_{WT} in this study and of its mutant defined as k_{mut} have been extensively measured by the kinetic isotope experiment (7, 10, 11, 21-32). Quantitatively, how a point mutation affect the hydride transfer rate of the DHFR can be described by $\Delta\Delta G = -RTln(\Gamma)$, the change in free energy difference for the hydride transfer reaction between a mutant and wild type, where $\Gamma = k_{WT}/k_{mut}$ is the fold of hydride transfer rate reduction and $RT = 0.6$ kcal/mol. As can be seen in table 1, large Γ means that the hydride-transfer rate is suppressed by the mutation, while near unity means that the rate stays unchanged.

Suppose that the mutation of amino acid will cause the intrinsic dynamics network breaking which further result in chemical modifications that allosterically regulate the enzyme activity - the hydride transfer rate reduction. To verify the hypothesis, we estimated the correlation between the CS of each sites and the corresponding averaged change in free energy difference

$\langle \Delta\Delta G \rangle$, where $\langle ... \rangle$ denotes the average over different mutation types of the same site. The results are reported in section 3e.

RESULTS

3a. The signal propagation and characteristic time

Shows movie of DHFR in Supporting Information.

Fig. 2(a) shows the relaxation motion of atom *i* (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 as $(\theta_j, \phi_j) = (0,0)$; 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. 2(c). The linear regression of $\tilde{t}_c^i(d_{ij})$ labeled by blue line give the propagation speed $\sim 580\ m/s$ with the correlation coefficient 0.76. Generally, the data shows that the atom *i* with larger $\tilde{t}_c^i$ is farther 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 regulate how fast could a perturbation propagate through the whole protein. Another interesting question is how fast the signal can propagate through the whole enzyme for a given impulse force $\vec{k}_j$ 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. Four groups in NADPH and three groups in DHF are classification (Fig. S2). For each residue *j*, a force ensemble with 133 directions was tried, then the slowest $t_c^i(\theta_j, \phi_j)$ terms for the longest dissemination time (LDT) of the residue *j*. The residues with short LDT are called the "disseminators". Figure 3 shows the LDT of all residues, where the LDT larger than 80 ps were neglected. It was found that several active sites (ILE 5, Phe 31, ILE 94, C6 of NADPH and N5 of DHF) are very good disseminators. The results suggest that the active sites locate at the optimal position of network to exploit the mechanical signal efficiently.

3c. The Coarse-Grain Connection Matrix (CGCM) and the Communication score

Following the method in section 2d, the CGCM are shown in Figure 4(a). There are several hot spots in the off-diagonal elements, especially, the first two highest elements, C(13,122)=1.52 and C(13, 121)=1.42 are the donor-acceptor pairs of V13-D122 and V13-G121, which turned out that the mutation of D122 and G121 can significant obstructed the hydride transfer rate. Interestingly, as shown in Table S3, we decomposed C(13,121) to 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 of signals than 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 5(a) 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. Comparing to the experimental mutation data that provide the ratio of wide-type hydride transfer rate (7, 10, 11, 21-26, 28) to that of the mutant, Γ, it is found that once those residues with high CSs were mutated, the hydride transfer rates of DHFR would significantly drop (resulting in a very small ρ value).

3d. The conservation of residues of DHFR in evolution

Further, we verify the evolutional importance of the identified ICCs. The number histogram of the ConSurf scores of all residues in DHFR is shown in Figure 5 with 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. We then calculated histograms of ConSurf scores with CSs higher than 0.9 (84 residues), 1.0 (57 residues), 1.1 (30 residues), 1.2 (16 residues) and 1.3 (5 residues), respectively. It can be seen that from the blue line (CS >0.9) to the red line (CS>1.3) that generally the higher the CSs of residues are, the more conserved these residues are (by ConSurf scores). The results support the aspect that dynamically significant residues are also evolutionally important (66).

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

By collecting experimental data from literatures, in the Table 1**,** we summarized wild-type amino acids to be mutated in the horizontal direction and the after mutation residues with alphabet codes in the vertical direction. The fold of hydride transfer rate reduction for a single point mutation is expressed as $\Gamma = k_{WT}/k_{mut}$. For example, the I40A mutation causes 40 fold of hydride transfer rate reduction. The red and blue highlights indicate active sites and perturbed sites, respectively. $\langle \Delta\Delta \mathrm{G} \rangle = \langle \Delta \mathrm{G}_{WT} - \Delta \mathrm{G}_{mutant} \rangle = \langle -\mathrm{RT} ln(\Gamma) \rangle$ is the average of a set of changes in free energy difference for the hydride transfer reaction between a wild type and its mutants where $\langle ... \rangle$ denotes the average of mutation types of a site. Figure 6 shows the negative $\langle \Delta\Delta G \rangle$ against the communication score, where the linear regression of 17 experimental data points gives the correlation coefficient of 0.8. The strong linear dependence suggests that the dynamical signals propagation of a residue (the maximum dynamical signals accessibility of a residue) play an important role of the allostery of the hydride transfer reaction.

DISCUSSIONS

There have been evidence from both experiments [] and simulations [] providing insights in mechanistic details on enzyme catalysis driven by intramolecular networks. Although the correlation between the allosteric sites and the ICCs is found in our study, questions still remain as for how this “communication” property is related to known basic geometric or physiochemical features in the enzyme. It is not immediately clear at the moment. An 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; Yang and Bahar, 2005), nor are they located at the peaks of highest frequency modes (indicating local folding core Bahar et al., 1998) (Fig S5). Several factors have been proposed to affect the intramolecular dynamics propagation network: hydrogen bonds within the local motif [], the contact number of residues [], the flexibility of the local motif []. Furthermore, beyond the static structure properties, the characters of vibrational modes such as: mode-coupling and localization [] and solvent friction [] should also be considered. Not all of the factors are independent. There are several important hydrogen bonds D122-G15, D122-E17 and S148-N23 corresponding to stabilize the F-G loop and F-G loop, respectively. (See Fig.S1)

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 for the hydride transfer reaction between a wild type with its single mutants of both residues and with a double mutant, $\Delta\Delta G = RTln(\Gamma)$ (11, 28), where $\Gamma = k_{WT}/k_{mut}$, is the fold of the reduction of hydride transfer rate for a single point mutation or a double mutation. It was proposed that a degree of coupling of two distal residues can be measured by the nonadditive property of coupling energy ΔG_I (11), where the coupling energy ΔG_I between two residue A and B is $\Delta G_I = \Delta\Delta G_{A,B} - (\Delta\Delta G_A + \Delta\Delta G_B)$. The larger the magnitude of coupling energy $|\Delta G_I|$, the higher degree of coupling between the two selected residues is. For example, given Γ for three single mutants M42W$(\Gamma = 41)$, G121V$(\Gamma = 163)$ and F125M$(\Gamma = 42)$, the coupling energies $|\Delta G_I|$ of three double mutants: M42W-G121V ($\Gamma = 7600$), G121V-F125M ($\Gamma = 61$) and M42W-F125M ($\Gamma = 512$) (28) are 0.077, 2.83 and 0.73, respectively. According to the coupling energies, the results indicate that M42W-G121V is the weakest coupling pair with the smallest $|\Delta G_I| = 0.077$. On the other hand, the strongest coupling pair is G121V-F125M with the largest $|\Delta G_I| = 2.83$. For comparison, we look for the corresponding elements of donor-acceptor pairs of the CGCM, where C(42,121)=0.26 and C(121,42)=0.24, C(121,125)=0.59 and C(125,121)=0.66 and C(42,125)=0.46 and C(125,42)=0.45. Based on the definition of CGCM, the larger the elements of CGCM, the more signals propagate between the two residues. Our results shows the same tendency of coupling strength G121V-F125M > M42-F125M > M42W-G121V with the experiments, and suggests the counts of signals between residues are also relating to the couplings strength between residues.

The robostness of CS

We have test several sets of perturbation sites, from p6, p7, p24 to p159. Overall the conservation dependence (Fig 5) was not change (change a little bit). The correlation between CS and experimental mutants data varies slightly (Fig 5a, 5b)

In addition to the local structure dynamics such as the flexibility predicted by NMR [] or contact number from the structure information [], the CSs provide a quantitative measurement describing “the probability 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 facilitating the enzyme catalysis [], also provides a solid theoretical basis beyond previous evidence [NMR, bioinformatics, MD].

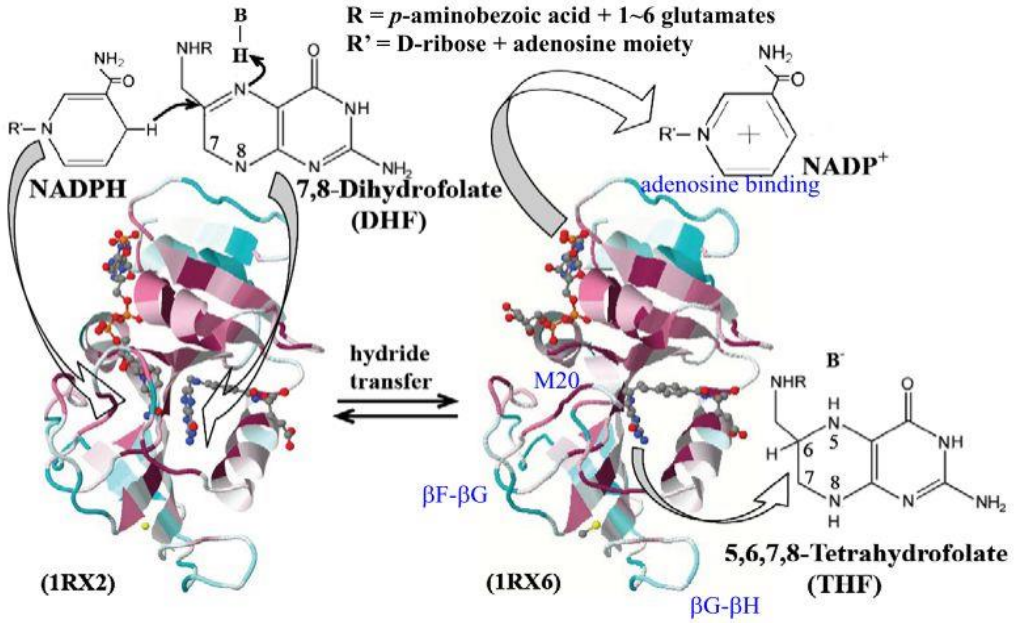

Figure 1 E. coli DHFR (simply DHFR therefore) facilitates the hydride transfer to C6 of the pterin nucleus with concurrent proton-ation in N5 position of DHF in the reaction $\mathbf{NADPH + H^+ \rightarrow NADP^+ + THF}$. The hydridetransfer-prompting Michaelis complex on the left (PDB ID: 1RX2) and the product ternary complex on the right (PDB ID:1RX6) are color-coded by residue conservation based on multiple sequence alignment of 77 DHFR homologs from cyan, the most variable, to dark red, the most conserved.

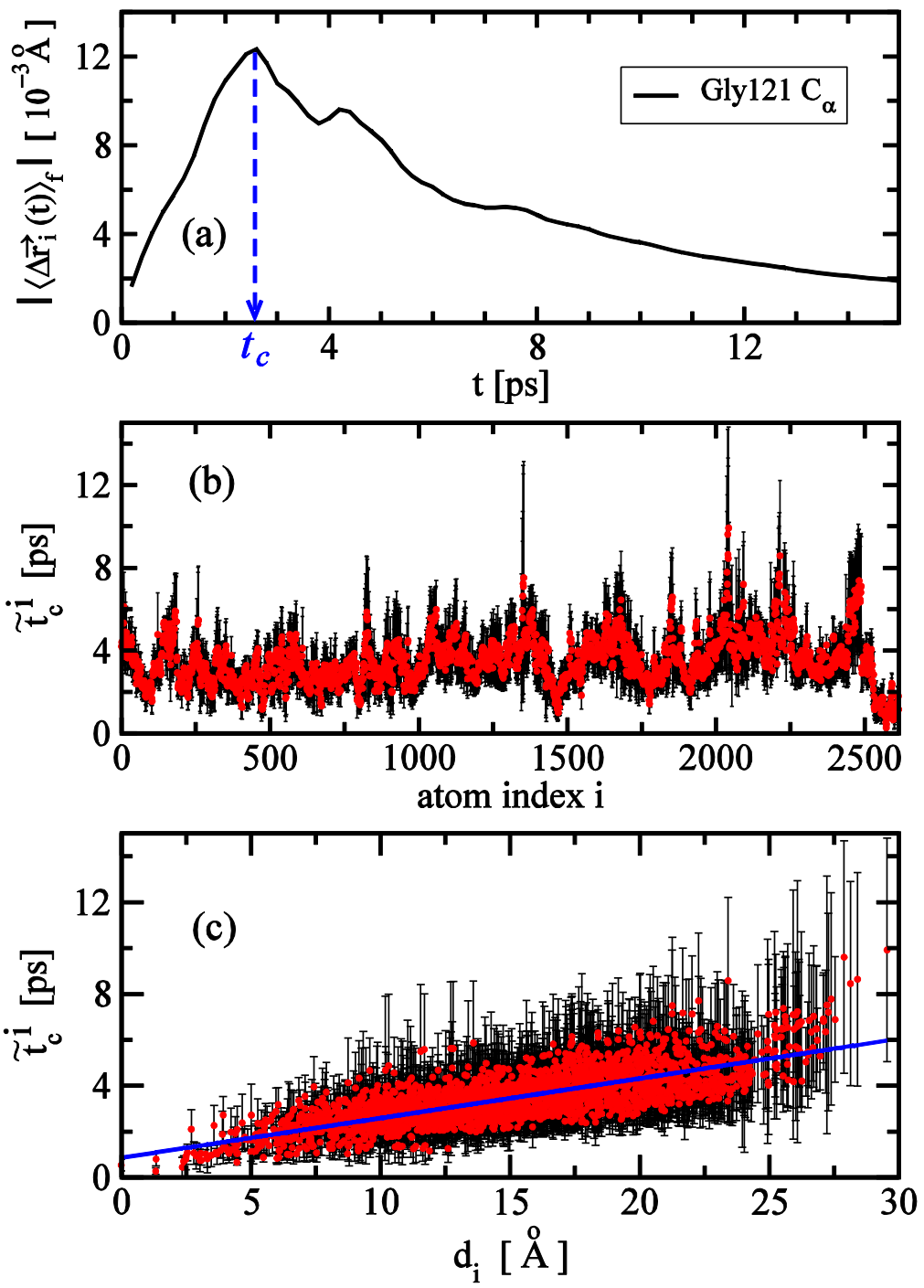

Figure 2 (a) The displacement of Gly 121 C_α atom against the response time with the characteristic time t_c=2.6 ps predicted by LRT using an impulse force on the N5 atom of the DHF. (b) By applying 133 evenly distributed forces on the N5 atom of the DHF, a red dot is the average characteristic times $\tilde{t}_c^i$ 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 0.76. The inverse slope provided to estimate the propagation speed of $580\ m/s$, and the estimated speed is compatible to results of a few theoretical groups (41, 45).

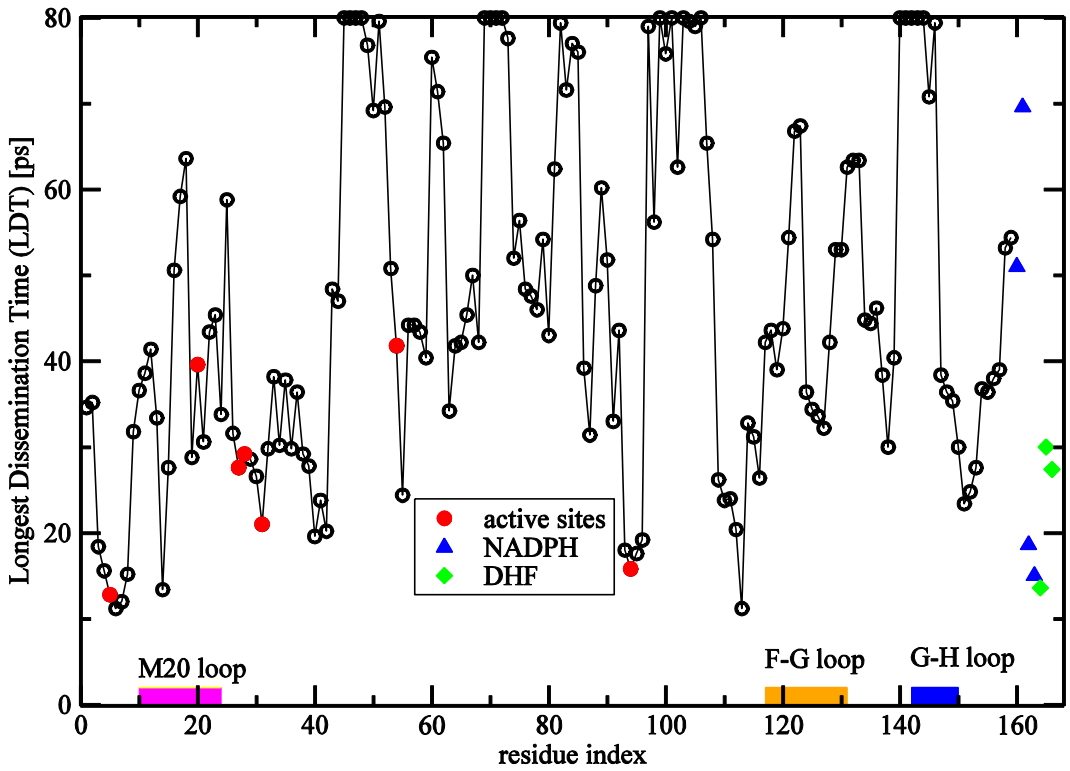

Figure 3 The longest dissemination times (LDTs) of selected indices of sites. An LDT is the slowest response time defined by the longest t_c in the DHFR with 133 evenly distributed impulse forces starting from a C_α atom of a selected residue. Indices 160-163 and 164-166 indicate 4 groups and 3 groups of NADPH and DHF, respectively. 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 small LDTs. For our studying purpose, we do not care about large LDTs, thus, an estimation of an LDT over 80 ps is shown as 80 ps. The solid red circles indicate the active sites: Ile5, Met20, Asp27, Leu28, Phe31, Leu54 and Ile94 (62).

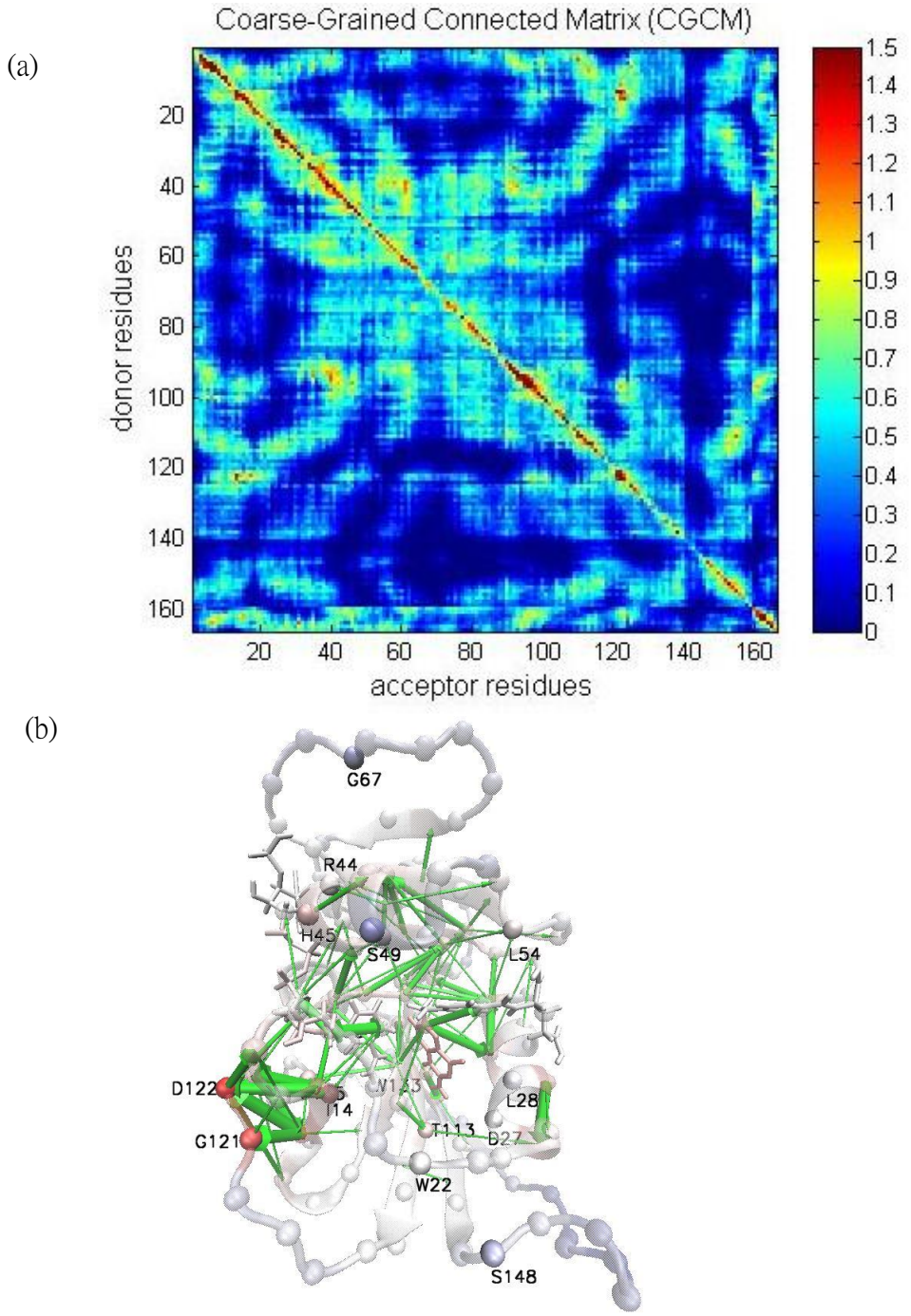

Figure 4 (a) The coarse-grained connection matrix (CGCM) in which each element, $C(R_d, R_r)$, records the counts of mechanical signals propagating between a donor (R_d)-accetpor (R_r) residue pair is generated by the method in section 2d. Consequently, a unique "communication score (CS)" is assigned to a selected residue with the highest CGCM off-diagonal element satisfying $|R_d - R_r| > 2$ corresponding to that residue (either be a donor or an acceptor). In another words, the CS of residue *i* is the highest CGCM score of the *i-th* row or column. For example, selecting the residue 122, as a donor, the highest CGCM 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 minimum CS to zero (dark blue) and the maximum CS to one (dark red). The green arrows pointing from donors to acceptors indicate signal propagation pathways, where the thickness represents the size of CSs.

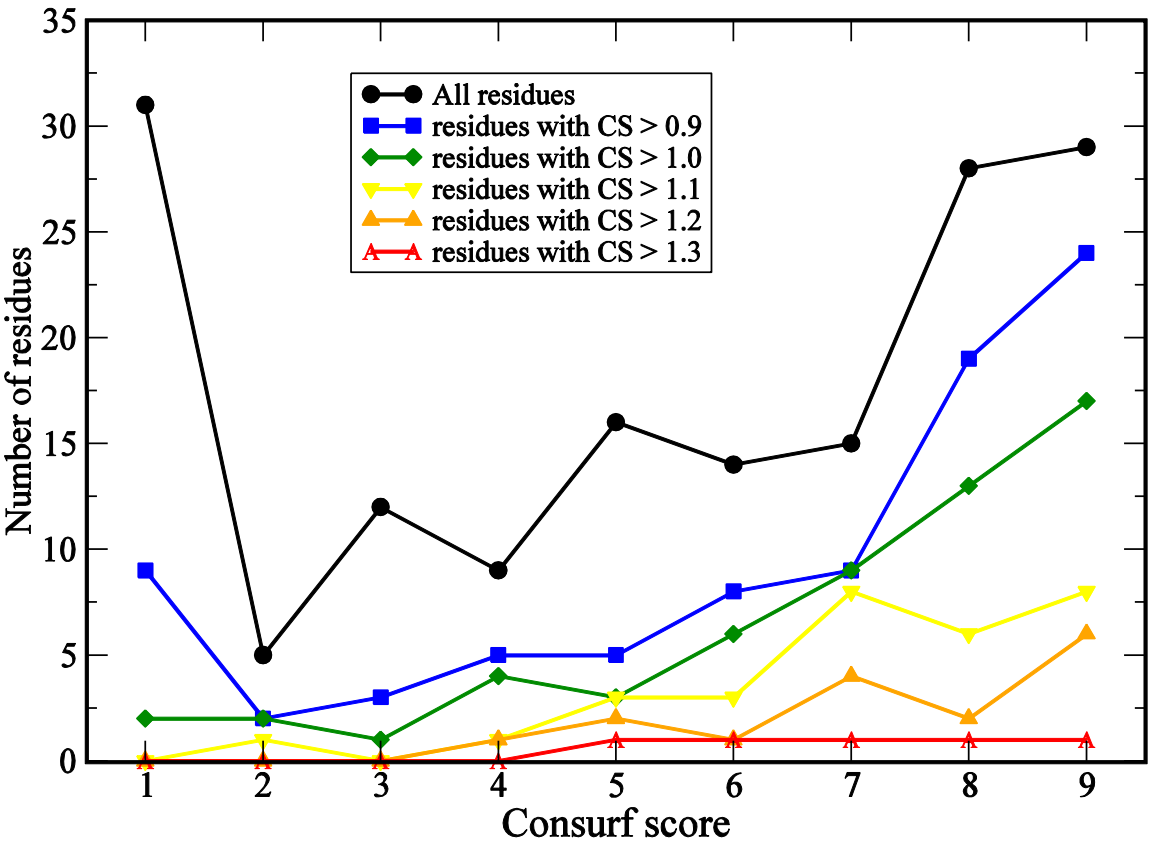

Figure 5 The histograms of ConSurf scores for the residues with different CSs, where the ConSurf scores representing the conservation of residues calculated from ConSurf web site (63-65). The higher the ConSurf score, the more conserved the residue is. The black line indicates the distribution for all 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. The fold of hydride transfer rate reduction, k_{WT}/k_{mut}, for mutants

WT / mutant	14	22	27	28	42	44	45	49	54	67	100	113	121	122	125	133	148
	I	**W**	**D**	**L**	**M**	**R**	**H**	**S**	**L**	**G**	**Y**	**T**	**G**	**D**	**F**	**W**	**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 collecting experimental data from literatures (see the References in the bottom row), we summarized wild-type (WT) amino acids in the top row and their corresponding mutants in the left column. The numbers indexed by mutant and WT amino acids are the fold reduction of hydride transfer rate for a single point mutation and is defined as $\Gamma = k_{WT}/k_{mut}$. For example, the I40A 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 perturbed 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 (the amino acid in the top row). CS is the communication score; $\mathbf{D_{act}}$ is the distance between the middle point of N5 and C6 atom of DHF (active site) and the C_α atom of the mutated site. The ConSurf score (63-65) represents the conservation of a residue, where score 9 is the most conserved and 1 is the least.

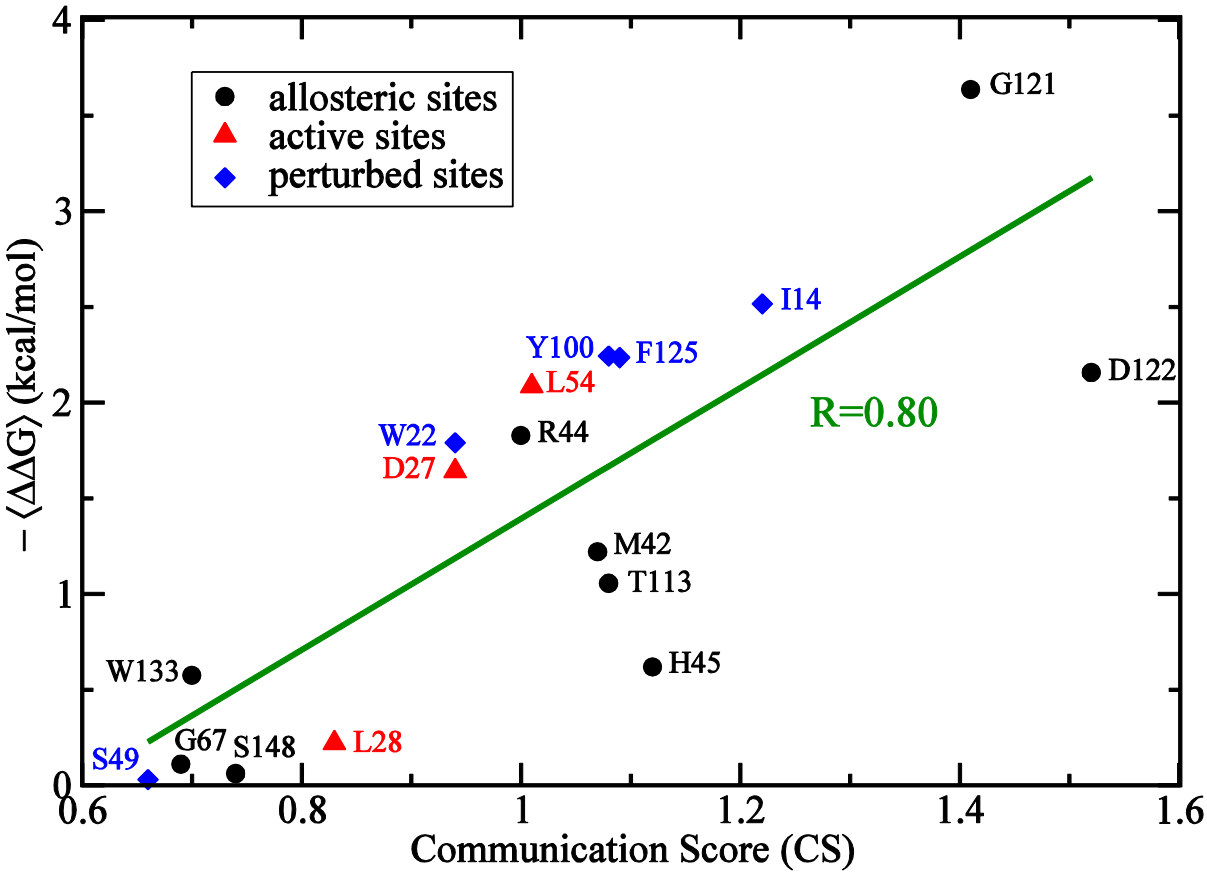

Figure 6 The averaged change in free energy difference for the hydride transfer reaction between a wild type and its mutants against the communication score (CS). The averaged $\langle \Delta\Delta \mathbf{G} \rangle = \langle \Delta \boldsymbol{G_{WT}} - \Delta \boldsymbol{G_{mutant}} \rangle = \langle -RTln(k_{WT}/k_{mut}) \rangle$, 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 mutation types of the same site; the data set is listed in Table 1. The CS of each residue is determined following the method in section 2d, and CS for all residues of DHFR is listed in Table S1. The red triangles are active sites, blue diamonds are perturbed sites and black circles are allosteric sites. The dark green line shows linear regression of total 17 data with the high 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.

54. Gekko, K., Y. Tamura, E. Ohmae, H. Hayashi, H. Kagamiyama, and H. Ueno. 1996. A large compressibility change of protein induced by a single amino acid substitution. Protein science : a publication of the Protein Society 5:542-545.

55. Gekko, K., and Y. Hasegawa. 1986. Compressibility-structure relationship of globular proteins. Biochemistry 25:6563-6571.

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. Celniker, G., G. Nimrod, H. Ashkenazy, F. Glaser, E. Martz, I. Mayrose, T. Pupko, and N. Ben-Tal. 2013. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. Israel Journal of Chemistry 53:199-206.

65. Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, and N. Ben-Tal. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic acids research 33:W299-302.

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

SUPPORTING INFORMATION

Movie of ICCs of DHFR

The active sites and perturbed sites

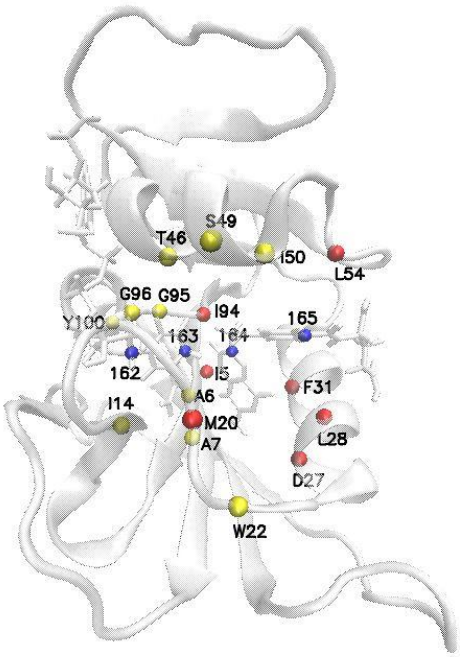

Figure S2. The 21 spheres are perturbed by impulse forces, where the red spheres are active sites: ILE5, MET20, ASP27, LEU28, PHE31, LEU54 and ILE94; yellow spheres are residues within 7 **Å** from the catalytic center: ALA6, ALA7, ILE14, TRP22, THR46, SER49, ILE50, GLY95, GLY96, TYR100. The four blue 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.

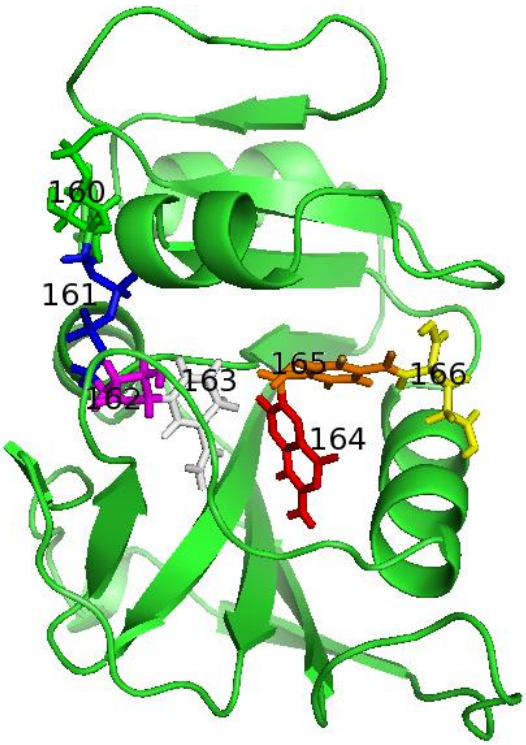

Figure S3. The coarse-grained NADPH and dihydrofolate. The NADPH is divided into four groups, 160-163, colored by green, blue, purple and white, respectively. The dihydrofolate is divided into three groups 164-166 colored by red, orange and yellow.

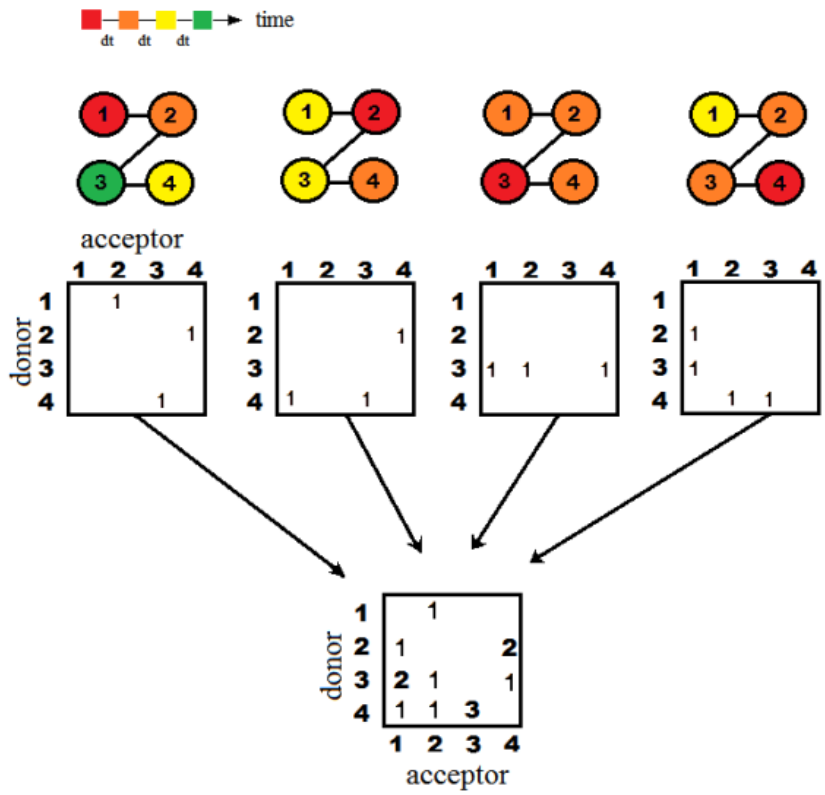

Figure S4 Recording the signals into the connection matrices. The colors indicate a time sequence of characteristic times. The circles with number linked by solid line are the model protein sequences, where the number is the atomic index. Given different perturbed forces, $\vec{k}_j$, the signals propagation 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 $F(a_d, a_r)$.

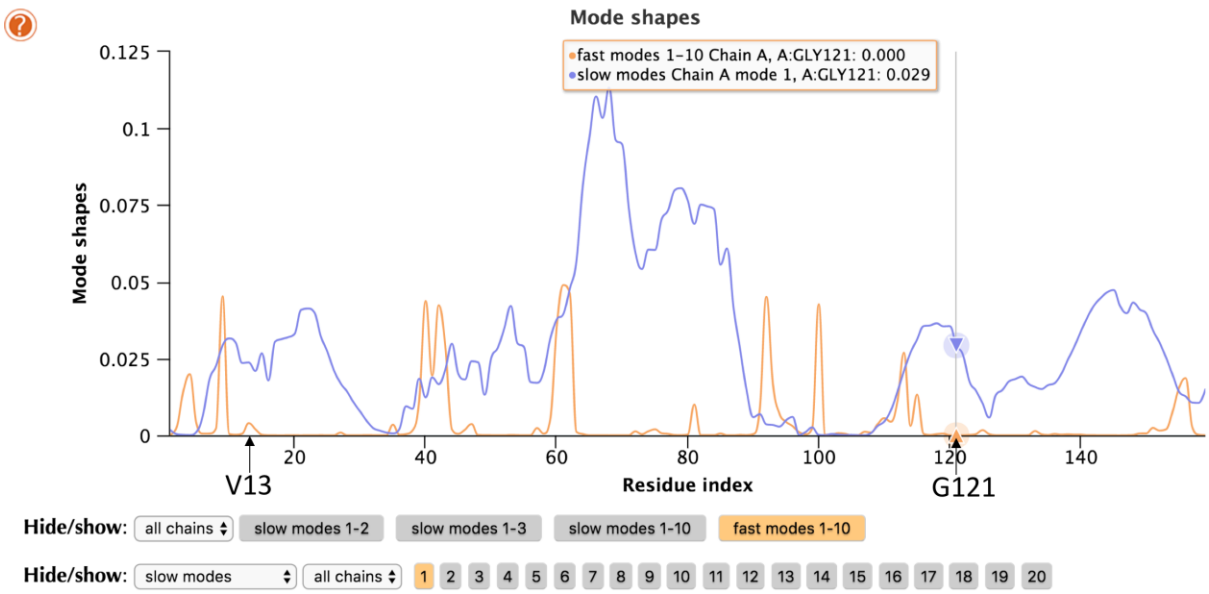

Figure S5 The ENM slowest mode (orange) and the average of the fastest 10 ENM mode (blue) for ternary DHFR complex (1RX6) are plotted against residue index. Frequent communicators V13-G121 are marked by black arrows. The analysis and resulting figure are taken from the DynOmics website (Li et al., 2017;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

rid / score		rid / score		rid / score		rid / score		rid / score		rid / score		rid / score		rid / 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 (rid).

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), 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

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

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.