Article

Ligand-Induced Protein Responses and Mechanical Signal Propagation Described by Linear Response Theories

Lee-Wei Yang,[1,*] Akio Kitao,[2,*] Bang-Chieh Huang,[1] and Nobuhiro Gō[3]
[1]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan; [2]Institute of Molecular and Cellular Biosciences, University of Tokyo, Tokyo, Japan; and [3]Quantum Beam Science Directorate, Japan Atomic Energy Research Agency, Kyoto, Japan

ABSTRACT In this study, a general linear response theory (LRT) is formulated to describe time-dependent and -independent protein conformational changes upon CO binding with myoglobin. Using the theory, we are able to monitor protein relaxation in two stages. The slower relaxation is found to occur from 4.4 to 81.2 picoseconds and the time constants characterized for a couple of aromatic residues agree with those observed by UV Resonance Raman (UVRR) spectrometry and time resolved x-ray crystallography. The faster "early responses", triggered as early as 400 femtoseconds, can be best described by the theory when impulse forces are used. The newly formulated theory describes the mechanical propagation following ligand-binding as a function of time, space and types of the perturbation forces. The "disseminators", defined as the residues that propagate signals throughout the molecule the fastest among all the residues in protein when perturbed, are found evolutionarily conserved and the mutations of which have been shown to largely change the CO rebinding kinetics in myoglobin.

INTRODUCTION

The flexible nature of proteins is crucial to their biological functions. Myoglobin (Mb), for instance, expressed solely in cardiac myocytes and oxidative skeletal muscle fibers, exhibits the feeble breathing motion in the central pocket that allows the binding and release of solvated gas molecules (1–4) and therefore secures our adaptability to fluctuations in blood oxygen levels.

Protein can reconfigure itself in a repertoire of possible ways, yet a particular conformational change in protein would take place in response to binding of specific ligands, which implies that the ligands select or trigger specific sets of motions (modes), possibly intrinsic to the proteins (1,5), to result in the observed conformational changes.

Recently, there has been renewed interest in using linear response theory (LRT) (2,6–9) to describe such ligand-triggered motions. In the framework of LRT, observed conformational changes result from a delicate interplay between protein intrinsic dynamics and the external perturbations that excite the most relevant global modes (2). Ikeguchi et al. showed excellent agreement between experimentally characterized and LRT-predicted conformational changes for three ligand-binding proteins (2). The molecular responses to ligand binding were described by the product of protein fluctuation covariance and the external forces applied by ligands (2).

The Coalson group also derived a time-dependent version of the theory to characterize fine- and coarse-grained presentations of the ligand binding process on timescales of hundreds of picoseconds or longer (6,7). The authors also showed that the exaggerated curvature might be overcome by quasiharmonic approximation of the roughness of the energy landscape caused by anharmonicity (7). Kidera and colleagues further demonstrated that LRT can well explain ligand-triggered domain motion for >50 cases in which ligands exert forces on multiple points of proteins that are within 4 Å of them (8).

In this study, we formulate two time-dependent LRTs that use constant and impulse forces to address the physical nature of two-stage time responses in Mb as shown by ultraviolet resonance Raman (UVRR) data. We demonstrate the predictive power of LRT on molecular relaxation occurring at a timescale from hundreds of femtoseconds to ~100 ps. With the new formula, in which impulse forces are used, mechanical signal propagation upon ligand binding can be tracked in terms of both time and space.

We further examine how fast perturbations on a specific residue can propagate throughout the molecule. The residues that most rapidly disseminate signals are found to be evolutionarily conserved. Evidence shows that mutation of these fast disseminators, including a few that are >9 Å away from the binding sites, alters the CO rebinding kinetics to various degrees. Our method is shown to predict the impact of these mutants on rebinding kinetics with good sensitivity.

THEORIES AND METHODS

The goal of this study is to derive a general theory that accounts for perturbation-induced protein conformational changes. The framework of the theory follows standard statistical mechanics (9,10). In this study, we specify the physical quantities in LRT such that the after-effect function is the atom

Submitted February 26, 2014, and accepted for publication July 15, 2014.
*Correspondence: lwyang@life.nthu.edu.tw or kitao@iam.u-tokyo.ac.jp
Editor: Nathan Baker.

0006-3495/14/09/1415/11 $2.00
http://dx.doi.org/10.1016/j.bpj.2014.07.049

fluctuation covariance and the external field is the perturbation forces (see below and Supporting Material). The acquired formulation is time-dependent. The time-independent form of LRT is deduced from the time-dependent form when the time-correlation term vanishes at time infinity (see below).

A general theory: time-dependent LRT (td-LRT)

H_o and H_f are the Hamiltonian terms for unperturbed and perturbed systems, respectively. The perturbed system is subject to forces $\mathbf{f}_j$ applied on atom j ($j = 1$ to m) such that $H_f = H_o - \sum_{j=1}^{m} \Delta\vec{\mathbf{r}}_j \mathbf{f}_j$, where $\Delta\vec{\mathbf{r}}_j$ is the deviation (from the mean) of atom j due to the external force, $\mathbf{f}_j$. $\langle \Delta\vec{\mathbf{r}}_i(t) \rangle_f$, the time progression of the positional changes of atom i under external forces (noted by subscript f), is derived (see Supporting Material) and assumes the form

$$\langle \Delta\vec{\mathbf{r}}_i(t) \rangle_f = \frac{1}{k_B T} \int_0^t dt' \sum \left\langle \Delta\dot{\vec{\mathbf{r}}}_j(0) \Delta\vec{\mathbf{r}}_i(t') \right\rangle_0 \mathbf{f}_j(t - t'), \quad (1)$$

where $\Delta\dot{\vec{\mathbf{r}}}_j(0)$ is the velocity of atom j at the moment when external forces, $\mathbf{f}_j$, are applied. $\langle \Delta\dot{\vec{\mathbf{r}}}_j(0) \Delta\vec{\mathbf{r}}_i(t') \rangle_0$ is the velocity-position time-correlation function sampled in the absence of perturbations (noted by subscript 0). The functions 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 (11) (see Supporting Material). k_B and T are the Boltzmann constant and absolute temperature, respectively.

td-LRT with constant force

If $\mathbf{f}_j$ is a constant force and the stationary condition $d/ds \langle \Delta\vec{\mathbf{r}}_i(s) \Delta\vec{\mathbf{r}}_j (s + t) \rangle_0 = 0$ can be applied, the above equation reduces to the form

$$\langle \Delta\vec{\mathbf{r}}_i(t) \rangle_f = \frac{1}{k_B T} \sum \left(\left\langle \Delta\vec{\mathbf{r}}_j(0) \Delta\vec{\mathbf{r}}_i(0) \right\rangle_0 - \left\langle \Delta\vec{\mathbf{r}}_j(0) \Delta\vec{\mathbf{r}}_i(t) \right\rangle_0 \right) \mathbf{f}_j. \quad (2)$$

Equation 2 is used to generate the time-resolved trajectory of the perturbed system. In this study, the perturbed system is CO-bound myoglobin (Mb-CO). The second term vanishes at time infinity, which leads to the time-independent form of LRT (Eq. 3). Thus, the current formula differs from the treatment of Essiz and Coalson (7) in a few respects. First, the derived results contain a time-independent term, and second, the formula employs the Langevin equation as treated by Chandrasekhar, including Gaussian random forces and solvent friction based on the approach of Hayward and colleagues (12), whereas Essiz et al. use the Stokes law to handle solvent friction (7).

The term $\langle \Delta\vec{\mathbf{r}}_j(0) \Delta\vec{\mathbf{r}}_i(t) \rangle_0$ can be expressed in the normal-mode space such that $\langle \Delta\vec{\mathbf{r}}_j(0) \Delta\vec{\mathbf{r}}_i(t) \rangle_0 = \sum_m \boldsymbol{\alpha}_{jm} \boldsymbol{\alpha}_{im} \langle \sigma_m(0)\, \sigma_m(t) \rangle_0$, where σ_m is the normal-mode variable (12–14) of mode m and $\boldsymbol{\alpha}_{jm}$ are the eigenvector components of atom j in that mode. The variables $\boldsymbol{\alpha}_{jm}$, $\boldsymbol{\alpha}_{im}$, and $\sigma_m(0)$ are obtained by normal-mode analysis (NMA) (see the Supporting Material). The term $\langle \sigma_m(0)\, \sigma_m(t) \rangle_0$ is readily derivable by solving the Langevin equation (11,12,15) where the multiple 1-D harmonic oscillators (normal-mode variables) are damped in the solvent and subject to random forces (for details, see the Supporting Material).

Time-independent LRT

When $t \to \infty$, the term $\langle \Delta\vec{\mathbf{r}}_j(0)\, \Delta\vec{\mathbf{r}}_i(t) \rangle_0$ vanishes and gives the time-independent form of the LRT (ti-LRT) (2), that is,

$$\langle \Delta\vec{\mathbf{r}}_i \rangle_f = \frac{1}{k_B T} \sum_j \left\langle \Delta\vec{\mathbf{r}}_j \Delta\vec{\mathbf{r}}_i \right\rangle_0 \mathbf{f}_j. \quad (3)$$

The above expression can be alternatively derived as described in the Supporting Material and in the work of Ikeguchi and co-workers (2). The constant forces herein can be derived using the same equation, such that $\mathbf{f}_j = k_B T \sum_i \langle \Delta\vec{\mathbf{r}}_j \Delta\vec{\mathbf{r}}_i \rangle_0^{-1} \langle \Delta\vec{\mathbf{r}}_i \rangle_f$ in kcal/mol/Å.

td-LRT using impulse forces

When $\mathbf{f}_j$ in Eq. 1 is a delta function such that $\mathbf{f}_j = \mathbf{k}_j \delta(t' - t'')$, where $t' \le t'' \le t$ (a pulse occurs at $t' = t''$), we obtain

$$\langle \Delta\vec{\mathbf{r}}_i(t) \rangle_f = \frac{1}{k_B T} \sum \left\langle \Delta\dot{\vec{\mathbf{r}}}_j(0) \Delta\vec{\mathbf{r}}_i(t) \right\rangle_0 \mathbf{k}_j, \quad (4)$$

where $\Delta\dot{\vec{\mathbf{r}}}_j(0)$ is the velocity of atom j at time zero. In this study, $\mathbf{k}_j$ are the forces, $\mathbf{f}_j$, derived from Eq. 3 multiplied by a small time interval. $\langle \Delta\dot{\vec{\mathbf{r}}}_j(0)\, \Delta\vec{\mathbf{r}}_i(t) \rangle_0$ is treated as damped harmonic oscillators(see Eqs. S5-6 and S5-7 in the Supporting Material). The formula is used to study the fast, early protein responses in the onset of CO binding as well as the mechanical signal propagation within the protein (see Results).

NMA

NMA (13,14) and solvent friction β calculation (12,15,16) are done according to established approaches. See the Supporting Material for details.

Energy minimization and calculation of structural differences

There is a single-residue difference between the deoxy- and CO-bound x-ray structures (PDB codes 1A6N and 1A6G, respectively) (17). Residue position 122 in 1A6G is an asparagine, whereas that residue in 1A6N is an aspartic acid. We first did in silico mutation to replace the $N_\delta H_2$ of the Asparagine with an O in 1A6G. We then carry out separate energy-minimization procedures for the two structures, 1A6N (1247 heavy atoms, or 2516 atoms total, including hydrogen) and 1A6G_N122D (2516 + CO atoms), for 200,000 steps by conjugate gradient and line-search algorithms implemented in the package nanoscale molecular dynamics (NAMD) (18).

Thus, energy-minimized 1A6N was used to perform NMA as described previously. On the other hand, we best-fit (19) all the atoms of the energy-minimized 1A6G_N122D (we call this conformation CO_min1) except the CO atoms to the corresponding atoms of 1A6N, keeping track of the translation vector and rotation matrix obtained in the fitting process (19,20), which are later used to translate and rotate the CO ligand into the 1A6N. The best-fit CO and energy-minimized 1A6N form a new chimeric CO-bound structure that is further energy-minimized for an additional 200,000 steps. We term this final structure CO_final, and it is in turn best-fit to energy-minimized 1A6N for all the atoms except CO. CO_final is found to have a root mean-square deviation (RMSD) from the energy-minimized 1A6N of 0.25 Å, which is close to the reported RMSD of 0.29 Å between the two x-ray structures (1A6N and 1A6G) (16). In this article, the structural change from deoxy- to CO-bound is the structural difference between energy-minimized 1A6N and CO_final. CO_final is also referred to as Mb-CO throughout the text.

Reaction coordinates to probe time response of a given atom or residue in Mb

Reaction coordinate φ, used to probe the time response of a given atom, is defined as $\phi(t)_i = \|\langle \Delta\vec{\mathbf{r}}_i(t) \rangle_f\| / \|\langle \Delta\vec{\mathbf{r}}_i(\infty) \rangle_f\|$, where $\langle \Delta\vec{\mathbf{r}}_i(\infty) \rangle_f$ is the positional changes of atom i from the deoxy- to the CO-bound state, $\langle \Delta\vec{\mathbf{r}}_i(t) \rangle_f$ is the deviation vector pointing from atom i at the deoxy state to its position at instant t, and $\langle \Delta\vec{\mathbf{r}}_i(0) \rangle_f = 0$. The reaction coordinate of a given residue or a

secondary structural fragment is defined as $\Phi(t)_a = \sum_{i=1}^{n_a} \phi(t)_i/n_a$; i runs on all the n_a heavy atoms of the residue or the secondary structure, a.

For the early responses examined by impulse forces, we define the culminating time, t_c, as the time at which an atom or a given group of atoms reaches its maximal response, $\max(\|\langle\Delta\vec{\mathbf{r}}_a(t)\rangle_f\|)$, the maximal departure of atom a (or residue/helix a) from the unperturbed position. When only one atom is perturbed, it can be easily understood from Eq. 4 and Eq. S5-7 in the Supporting Material that changes in the size of a point impulse force (k_j) or exertion of such a force in the opposite direction ($-k_j$) would not change the t_c for any atom in the system.

Reaction coordinates to probe time response for a pair of atoms or residues in Mb

For a given atom pair i and j, the reaction coordinate ψ_{ij} for the pair is defined as $\psi_{ij}(t) = \|\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f\| - \min(\|\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f\|)/\max(\|\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f\|) - \min(\|\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f\|)$, where $\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f$ is the positional difference between atoms i and j, starting at $\langle\vec{\mathbf{r}}_{ij}(0)\rangle_f$ and fluctuating between $\min(\|\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f\|)$ and $\max(\|\langle\vec{\mathbf{r}}_{ij}(t)\rangle_f\|)$, which are the minimal and maximal linear departure, respectively, between atoms i and j found in the time course from zero to infinity. For residue pair a and b, the reaction coordinate Ψ_{ab} is defined as $\Psi_{ab}(t) = \|l_{ab}(t)\| - \min(\|l_{ab}(t)\|)/\max(\|l_{ab}(t)\|) - \min(\|l_{ab}(t)\|)$, where the difference vector for the geometric centers of a and b read as $l_{ab}(t) = \sum_{i\subset b}^{n_b}\langle\vec{\mathbf{r}}_i(t)\rangle/n_b - \sum_{j\subset a}^{n_a}\langle\vec{\mathbf{r}}_j(t)\rangle/n_a$, with n_a and n_b the number of heavy atoms in residues a and b, respectively.

Point perturbations on each residue in Mb to access its longest dissemination time

Unity impulse force is exerted on a selected residue i (at its C_α) in 21 evenly distributed directions, and the slowest response, defined by the longest t_c, in the molecule is acquired as the longest dissemination time (LDT) for residue i. For example, Val^{68} is perturbed with unity forces exerted in 21 different directions and the HγI1 atom on Val^1 is found to be slowest in the system to respond, with a t_c of 15.8 ps, when a unity force is exerted on the C_α of Val^{68} from the angles $(\theta, \phi) = (30°, 45°)$ (see Fig. 4, *inset*). We therefore find the LDT for Val^{68} to be 15.8 ps. As mentioned previously, it is found that when a point impulse force acts in totally opposite directions on atom j, the t_c for any other atom i (t_c^i) remains unchanged. Therefore, we define a force ensemble with the same magnitude pointing in 21 different directions as a function of orientational angles (θ, ϕ) in polar coordinates (see Fig. 4, *inset*), where θ is the angle between the direction of force and the z axis (*blue*) and ϕ is the angle between the projection vector of the force in the xy plane and the x axis (*red*). θ spans 0°–90° with a step size of 30°, and ϕ ranges from 0° to 315° with a step size of 45°, where dependent directions that give identical t_c^i are omitted (for more details, see the Supporting Material).

RESULTS AND DISCUSSION

The results mainly comprise three parts. First, we show that ultraviolet resonance Raman spectroscopy (UVRR)-characterized two-stage relaxation in Mb can be well explained by our td-LRT formula using constant and impulse forces. Second, we define the culminating time, t_c, to assess when the maximal early response is reached for a given residue upon ligand binding. By obtaining t_c for every residue, we can describe the molecular signal propagation for a given perturbation. Third, we apply point perturbation in 21 different directions on every residue and then trace when the signals can propagate throughout the molecule. Then, we examine the evolutionary conservation and kinetic properties for those residues that disseminate the signals the fastest.

For the first part of the two-stage relaxation, the aforementioned theories are applied as follows. We first take the x-ray structures of CO-bound and unbound myoglobins. $\langle\Delta\vec{\mathbf{r}}_i\rangle_f$ is the $3N$-d vector that is the structural difference between energy-minimized CO-bound and unbound structures (see details in Theories and Methods). Covariance matrix $\langle\Delta\vec{\mathbf{r}}_j(0)\,\Delta\vec{\mathbf{r}}_i(0)\rangle_0$ is constructed by NMA using the CHARMM22 force field for the unbound Mb. Here, the subscript 0 denotes that the myoglobin is in the unbound state. We then use Eq. 3 to obtain the constant forces, $\mathbf{f}_j$, that drive the observed conformational changes. Next, we calculate the term $\langle\Delta\vec{\mathbf{r}}_j(0)\Delta\vec{\mathbf{r}}_i(t)\rangle_0$ in Eq. 2 by considering normal-mode variables subject to solvent friction and random forces (8) (see the Supporting Material). Together with the obtained $\langle\Delta\vec{\mathbf{r}}_j(0)\Delta\vec{\mathbf{r}}_i(0)\rangle_0$ and $\langle\Delta\vec{\mathbf{r}}_i\rangle_f$, we can use Eq. 2 to obtain a time trajectory of the conformational changes, $\langle\Delta\vec{\mathbf{r}}_i(t)\rangle_f$ (see results in Fig. 2, *a–c*).

Ligand-induced forces that drive the Mb → Mb-CO conformational change are localized

The conformational change, $\langle\Delta\vec{\mathbf{r}}_i\rangle_f$, is obtained from energy-minimized and best-fit (19,20) Mb and Mb-CO structures following the procedures described in Theories and Methods. The biggest conformational changes are found at the binding pocket. With mean RMSD >0.6 Å, L29, R45, H64, S92, H93, K96, H97, and the heme have the most pronounced positional changes (Fig. 1 *a*). These residues are mainly located on the tips of helices E and F and in the CD loop, indicated by the warm colors in the inset of Fig. 1 *a* (see also Fig. 3 for the locations of these secondary structural elements). The RMSD between equivalent heavy atoms in Mb and Mb-CO is found to be 0.25 Å, which is comparable with the reported 0.29 Å structural difference between the two crystalline structures 1A6N and 1A6G, which are solved in an identical environment (17).

The magnitude of internal forces on atoms, derived from Eq. 3, is plotted along the atom index in Fig. 1 *b*. In Eq. 3, NMA performed on energy-minimized Mb (PDB 1A6N) gives the fluctuation covariance of the protein. Notably, NMA-predicted atom fluctuations have a correlation of 0.59 with the Debye-Waller factors of equivalent atoms reported in the x-ray structure (1A6N) and a correlation of 0.65 with the RMSDs of equivalent atoms calculated from 12 NMR conformers of Mb (1MYF) (21) (for a definition of the Pearson correlation coefficient, see the Supporting Material). The obtained correlations are comparable with those found previously between theory and experiment when large sets of structures were examined (21,22). A correlation found between the positional change, $|\langle\Delta\vec{\mathbf{r}}_i\rangle_f|$, of atom i and the size of the force exerted on the atom for all

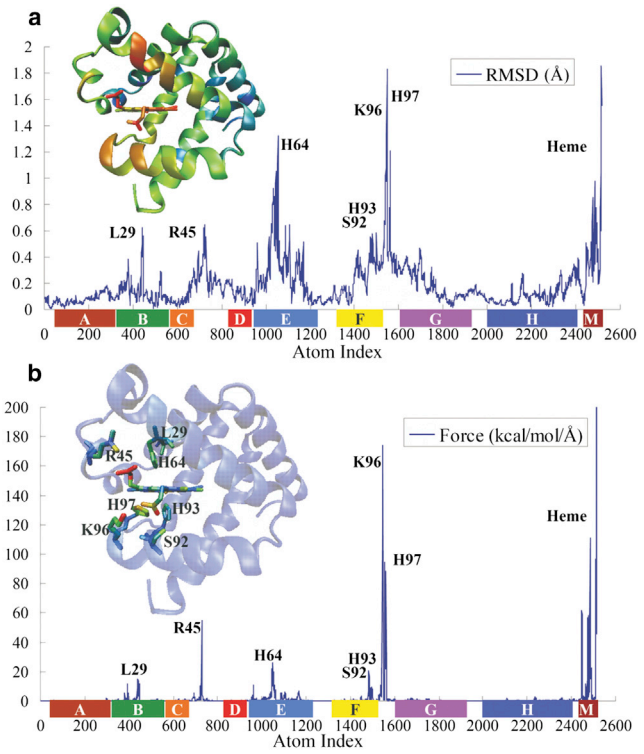

FIGURE 1 RMSD (Å) between deoxy-Mb (unbound) and Mb-CO (bound) structures (*a*) and (*b*) ti-LRT-derived forces (kcal/mol/Å) are plotted as functions of the atom index. The color bars lying under the abscissa denote the helices and the heme (M) in Mb. Under the color bar is the atom index (1247 heavy atoms and a total of 2516 atoms, including hydrogen) (see the Supporting Material for details). (*Insets*) Ribbon diagrams show atoms color-coded from blue to red in ascending order of the magnitude of structural difference (Mb→Mb-CO) (*a*) or force (*b*). In *b*, heme (≥2444 in abscissa) and the residues containing atoms under forces >15 kcal/mol/Å are plotted in licorice in the ribbon diagram.

the 2516 atoms, is 0.48. The moderate correlation means that generally stronger forces on atoms would result in pronounced positional changes, $\left|\langle\Delta\vec{\mathbf{r}}_i\rangle_f\right|$. Commensurate with the positional deviation results (Fig. 1 *a*), L29, R45, H64, S92, H93, K96, H97, and the heme are found to be subject to high forces of >15 kcal/mol/Å (~1 nN/molecule), whereas the average obtained for all atoms in Mb is 2.09 kcal/mol/Å (~140 pN/molecule). The locations of these residues are shown in licorice in the inset of Fig. 1 *b* color-coded blue→green→red in ascending order of force magnitude.

However, when not only the size but also the direction of conformational change is considered, the two $3N - d$ vectors, positional deviations, and forces (containing *x*-, *y*-, and *z*-components for each atom), have a normalized dot product of 0.07 (barely correlated; unity would mean perfect correlation). This implies that the conformational changes are not necessarily driven in the directions in which the forces are applied (otherwise, the theory would be trivial), despite the appreciable correlation with their magnitude. Note that the forces derived using Eq. 3 are internal, meaning that the forces and torques on all the degrees of freedom in the system sum up to zero.

Time responses described by td-LRT are in close agreement with UVRR relaxation data

As described in the Theory, we express the time-dependent and time-independent fluctuation covariance of atoms in the normal-mode space. The normal-mode approximation advantageously describes minute conformational changes in Mb upon CO binding. The use of Eq. 2 and Eqs. S4-4 and S4-5 in the Supporting Material to obtain time-resolved positional changes of atoms, $\langle\Delta\vec{\mathbf{r}}_i(t)\rangle_f$, requires the solvent friction coefficient and the frequencies/directions of normal modes. Following the protocol of Hayward and colleagues (12), we obtained the friction coefficient, β, of Mb in water as 27.6 cm^{-1} (Fig. S3; see details in the Supporting Material). The trajectory of the atoms is as shown in Fig. 2, *a* and *c*. The reaction coordinate $\Phi(t)_a$, which is a fraction of the size of destined positional deviations, Mb($t = 0$)→Mb-CO($t\rightarrow\infty$), is devised to outline the time response of residues or secondary structural elements of interest (see Theories and Methods). The relaxation process involving changes of atomic positions from the deoxy- to the CO-bound state is driven by the constant forces that can be derived from ti-LRT (Eq. 3).

Sato et al. previously reported ultrafast responses for residues Trp^{14} (on helix A) and Tyr^{146} (on helix H) in Mb (3). In Fig. 2 *a*, $\Phi(t)_a$ for Trp^{14} and Tyr^{146}, on helices A and H, respectively, is plotted as a function of response time, and these plots are readily compared with recently available UVRR data (3). The time-resolved spectrum monitors the picosecond changes of the hydrophobic/polarizable environment around tryptophan and tyrosine residues, dominantly attributed to Trp^{14} on helix A and Tyr^{146} on helix H, by tracking the time development of absorption intensities at the UV-characterized spectrum (3). The spectrum shows a sharp intensity change that occurs within 2 ps for both residues followed by a recovery process that spans a timeframe from 7.3–8.0 ps for Tyr^{146} to 45–59 ps for Trp^{14} (see Fig. S2).

When fitting the $\Phi(t)_a$ data to the exponential function $A[1 + B\exp(-t/\tau)]$ as used in the reported UVRR studies (3), from 4 ps to its saturation, we obtained time constants of $\tau =$ 21.6 and 10.7 ps for Tyr^{146} and Trp^{14}, respectively. The values appear inconsistent with the reported UVRR data, both qualitatively and quantitatively. As acknowledged by Sato et al. (3), the relaxation processes observed by UVRR result from the hydrophobicity changes encompassing the residues instead of simply from the departure of the residues from their own deoxy states. Ile^{99} was speculated to dominantly determine the hydrophobic environment of Tyr^{146}, because its backbone oxygen forms a hydrogen bond with the hydroxyl group of Tyr^{146} (3). On the other hand, Leu^{69} shows the closest nonlocal hydrophobic contact

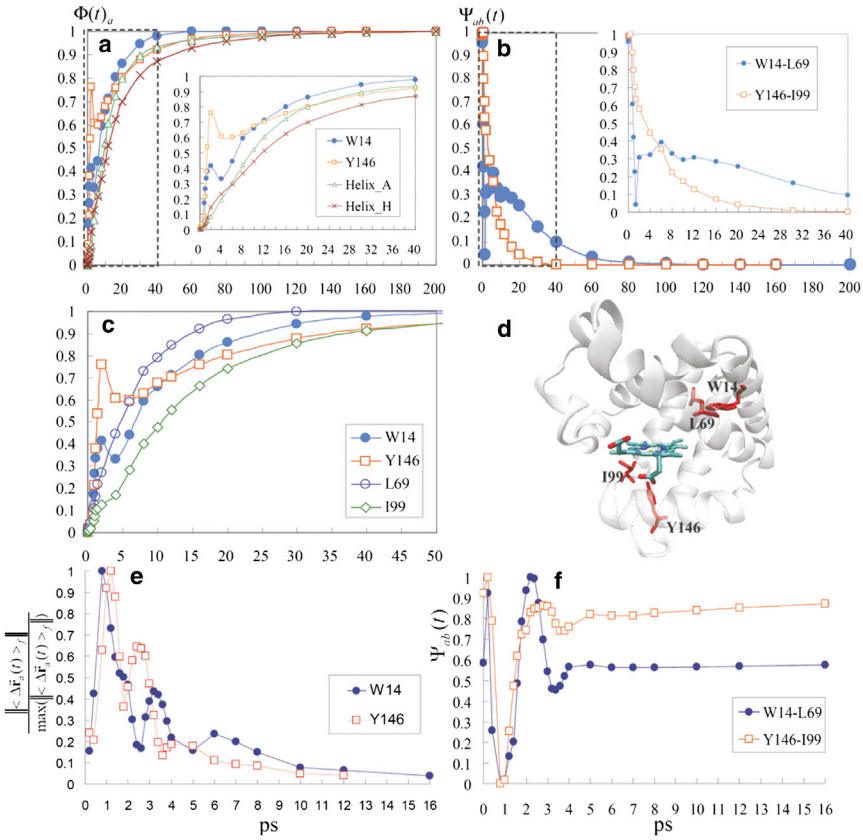

FIGURE 2 (*a*) The reaction coordinate $\Phi(t)_a$ for Trp^{14}, Tyr^{146}, and helices A and H are plotted as a function of response time. At time zero, constant forces are applied on the otherwise unperturbed molecule. Data points from 4 ps to the time at which Trp^{14} and Tyr^{146} reach 95% saturation are fitted to the exponential function $A[1 + B\exp(-t/\tau)]$, previously used by Sato et al. (3). The time constants, τ, obtained for Trp^{14} and Tyr^{146} are 10.7 and 21.6 ps, respectively. Data points before 4 ps are not considered in the fitting because of the early, fast responses culminating at 2 ps. (*b*) The normalized distance, $\Psi(t)_{ab}$ (see Methods), between Trp^{14} and Leu^{69}, as well as that between Tyr^{146} and Ile^{99}, are plotted against time. A minimal separation between residues *a* and *b* is reached when $\Psi(t)_{ab} = 0\%$, which happens to be the equilibrium separation for Trp^{14}-Leu^{69} and Tyr^{146}-Ile^{99}. Data points after 4 ps are fitted to the exponential function $B'\exp(-t/\tau')$. The time constants, τ', obtained for Trp^{14}-Leu^{69} and Tyr^{146}-Ile^{99} are 31.3 and 7.1 ps, respectively, comparable to those characterized by UV resonance Raman (see the Supporting Material). Data points before 4 ps are not taken into the fitting due to the nonexponential behavior (a sudden drop at 1.4 ps) of the Trp^{14}-Leu^{69} pair. Insets in *a* and *b* are the first 40 ps of the timeframe (*dashed outlines*). (*c*) Although L69 relaxes to the new Mb-CO state faster ($\tau = 6.0$ ps) than does W14, there is a noticeable constant difference between W14 and L69 for the timeframe 4–24 ps; whereas a diminishing difference between I99 ($\tau = 17.7$ ps) and Y146 is observed for the same timeframe. (*d*) Relative locations for Trp^{14}, L69, I99, and Tyr^{146} on the structure. (*e*) The fast primary responses for Trp^{14} and Tyr^{146} culminate at 800 fs and 1.2 ps, respectively. The values in ordinate have been normalized by dividing the maximal departure from the deoxy state so that the response culminates at unity. (*f*) In a similar way, the fast responses observed for Trp^{14}-Leu^{69} and Tyr^{146}-Ile^{99} have their closest contact distances at 800 fs before recovering to their original equilibrium states in an oscillatory fashion. This sudden drop and recovery occur on a timescale comparable to that characterized by UVRR (see discussion of τ_{decay} in the Supporting Material,).

with Trp^{14} (3), and it has been proposed previously that they jointly affect the binding kinetics (23).

The normalized relative distance change, $\Psi_{ab}(t)$, between Trp^{14} and Leu^{69} (W14-L69), as well as that between Tyr^{146} and Ile^{99} (Y146-I99), are shown in Fig. 2 *b*. Fitting the data to an exponential function $B'\exp(-t/\tau')$, from 4 ps to the time point when $\Psi_{ab}(t)$ drops to 5%, we obtain time constants τ' for Trp^{14}-Leu^{69} and Tyr^{146}-Ile^{99} as 31.3 and 7.1 ps, respectively, which well agree with the reported values 45–59 and 7.3–8.0 ps (see Fig. S2) (3). In the plot of $\Phi(t)_a$ responses for the aforementioned four residues together (Fig. 2 *c*), note the constant margin by which Leu^{69} ($\tau = 6.0$ ps) wins over Trp^{14} for the duration from 4 to 20+ ps approaching saturation. Over the same timeframe, a rapidly diminishing difference between Ile^{99} ($\tau = 17.7$ ps) and Tyr^{146} is observed. This may in part explain the delayed Trp^{14} relaxation that takes place near Leu^{69} (10.7→31.3 ps) and the expedited Tyr^{146} relaxation in the presence of Ile^{99} (21.6→7.1 ps).

Besides Ile^{99}, we find that Ala^{90}, forming a nonlocal hydrophobic contact (<5 Å) with Tyr^{146}, could also be responsible for the hydrophobic environment changes near Tyr^{146}. A time constant of $\tau' = 5.3$ ps is found for the normalized Tyr^{146}-Ala^{90} distance change (for further discussion, see the Supporting Material). These results demonstrate the predictability of td-LRT and further corroborate the early surmise of the importance of Leu^{69} and Ile^{99} in determining the intensity of the tryptophan (Trp^{14}) and tyrosine (Tyr^{146}) bands in UVRR spectra (3).

Femto- to picosecond early responses suggest mechanical signal propagation within Mb

In the analyses of the previous section, fitting to experimental data was taken after the fourth picosecond. This is due to distinctive sudden impulses in $\Phi(t)_a$ that occur at 2 ps for W14 and Y146 (Fig. 2 *a*, *inset*). Also, there is an instant drop of $\Psi_{ab}(t)$ for W14-L69 at 1.4 ps before its recovery (Fig. 2 *b*, *inset*). Interestingly, similar observations can be found in the published relaxation profile of T4 lysozyme (Figs. 2 *d* and 3 *c* in Essiz and Coalson (7)). The drastic changes have precluded certain data points from being incorporated into the aforementioned exponential fitting. We speculate that these quick responses are intrinsic

and irrelevant to the final conformation to which the deoxy-Mb relaxes. The assumption stems from the fact that the early responses reproducibly culminate at ~2 ps even when different end-state conformations, either CO_min1 (see definition in Theories and Methods) or x-ray structure 1A6G, are used. To further investigate the observed quick responses, we reformulated a td-LRT equation using impulse forces instead of constant forces as perturbation sources. By doing this, Mb relaxes to its original equilibrium state after a temporal reconfiguration in response to the perturbation (see Fig. 2 *e* and Eqs. S5-6 and 5-7 in the Supporting Material). Our focus here is the fast early responses that eventually dissipate in solvent friction as opposed to responses introduced in the previous section, which comprised both early responses and relaxation to a targeted conformation. The impulse forces used here are set to be proportional to the constant forces derived earlier (Fig. 1 *b*).

As shown in Fig. 2 *e*, the early responses for Trp^{14} and Tyr^{146} culminate at 800 fs and 1.2 ps, respectively. This timeframe is again observed in the early responses seen for Trp^{14}-Leu^{69} and Tyr^{146}-Ile^{99} pairs (Fig. 2 *f*). It can be seen that both pairs have their closest contact at 800 fs, followed by a recovery to their original separations in an oscillatory fashion. The identified timescale for the fast early response is comparable to that of ~2 ps characterized by UVRR (see discussion of τ_{decay} in the Supporting Material).

Let us define culminating time, t_c, as the time taken by an atom or a given group of atoms to reach a maximal response, $\max(\|\langle\Delta\vec{\mathbf{r}}_a(t)\rangle_f\|)$, given sets of impulse forces perturbing the system (see Theories and Methods). We then examine t_c for the heme, helices, loops, linkers, or individual residues in Mb. Residues with small t_c receive mechanical signals earlier than do those with large t_c. When examining such mechanical signal propagations by each of the secondary structural elements, we found that the heme and FG corner respond the fastest to the perturbations (within 400 fs (Fig. 3 *a*)), followed by the CD loop (1 ps), then F (1.2 ps), and then E + B helices (1.4 ps). G and H helices ($t_c = 1.6$ ps) respond twice as fast as A and C helices (3.2 ps) (see Fig. 3 *a*). The prediction that helix A is a slower responder than helix H agrees with the conclusion drawn by Sato et al. from single-residue evidence, namely, from Trp^{14} in helix A and Tyr^{146} in helix H (3). However, as the relaxation to the end-state conformation is considered, the τ values from exponential fitting found for helices A and H are 17.1 and 24.8 ps, respectively (Fig. 2 *a*), which shows that they relax at somewhat similar speeds.

When we examined signal propagation at the single-residue level (Fig. 3, *b* and *c*), we found that t_c occurs within 1 ps for W14, G25, **N26**, L29, I30, D44, R45, F46, S58, **E59**, D60, V66, **V68**, L69, L89, A90, **H93**, A94, **K96, H97**, I107, and **heme** (residues in bold print respond in <400 fs). They are mainly located in helices E and F, the FG-corner (K96 could have facilitated the signal propagation; see the first section in Results and the Supporting Material), a portion of the CD loop centered at R45 facing the heme, a few residues on helix B centered at L29, and the W14 on helix A (Fig. 3 *b*). Distal histidine His^{64} on the helix E is not among the fastest responders, but it too culminates quite rapidly, at $t_c = 1.4$ ps. Among the early responders, [L29, F46, H64, V68, and I107] form the heme cavity and gate CO release upon ligand dissociation (19). We found that those residues senseCO binding within 1–1.4 ps. On the other hand, residues with $t_c > 6$ ps appear to be far away from the heme, located either at the end of helix G (H116 and S117), on the GH linker (G121 and F123), at the beginning of helix H (G124, A125, D126, A127, and Q128), on the N-terminal residues (S3, E4, G5, and E6), on the EF linker (H81 and E83), or on the CD loop (T51) (Fig. 3 *c*). Among them, G5, F123 and G124 culminate the slowest, at 8 ps. The robustness of these data are further explored in the Results and Discussion section.

To test the robustness of our results, we apply a point force on the Fe ion pointing from the NE2 atom of the proximal histidine (H93) toward the Fe in the heme. Arbitrary magnitude is assigned, since t_c is not a function of the force magnitude for single-point-force cases. Such setup assumes that neither electrostatic nor van der Waals perturbations are imposed on spatially adjacent atoms near the CO. We see a systematic delay of t_c for those helices and secondary structural elements. The mechanical signal propagates in the order heme (800 fs) → helix F (1.6 ps) → FG corner (1.8 ps) → helix E (2.0 ps) → CD loop/helix G (2.2 ps) → helix B (2.4 ps) → helix D/helix A (2.8 ps) → helix C (3.2 ps) → helix H (5 ps) → EF linker (6 ps). Compared with the results in Fig. 3 *a*, the delay of t_c for the secondary elements is between 400 fs and 1.4 ps, except in the case of helix H, which has the longest delay, 3.4 ps. Despite the delay, the order of signal propagation is held unchanged, except that the order of helices G and B is switched, and helix H, an intermediate responder (1.6 ps) in Fig. 3, becomes a late responder (5 ps) in the current case of single-point force. This indicates that the originally faster response of helix H is mediated through the electrostatic influence of CO instead of through the CO-Fe-His^{93} bonding. To conclude, we found in our study that the fast, early responses are rather robust.

Mode-by-mode contributions to the equilibrium and time-resolved molecular responses

We examined the static contribution of modes to the $\langle\Delta\vec{\mathbf{r}}_i\rangle_f$ in Eq. 3. A handful of slow modes, the 2nd, 3rd, 8th, 11th, and 12th modes, with frequencies 5.6, 6.2, 8.4, 9.6, and 10.1 cm^{-1}, respectively, are found to be highly excited by the CO-imposed forces and to contribute >20% of the overall motional magnitude, $\|\langle\Delta\vec{\mathbf{r}}\rangle_f\|$ (see Fig. S4). Among those, mode 11 has the largest contribution (56%) to $\|\langle\Delta\vec{\mathbf{r}}\rangle_f\|$.

On the other hand, it is also interesting to see the mode contribution to the t_c. As impulse forces are used, the

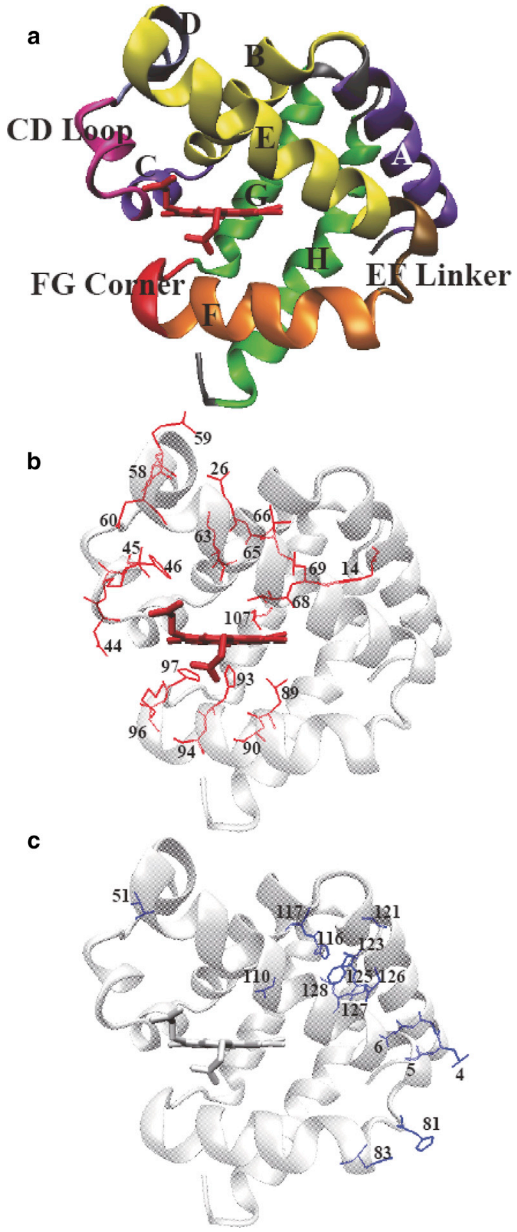

FIGURE 3 (*a*) Secondary structure elements, helices A–H, the CD loop, the EF linker and the FG corner (see Results for definitions), and heme are color-coded in ascending order of t_c defined according to the time response of the mean positions of the heavy atoms within a given segment, from red (400 fs (heme and FG corner)), magenta (1.0 ps (CD loop)), orange (1.2 ps (F helix)), yellow (1.4 ps (E and B helices)), green (1.6 ps (G and H helices)), ice blue (2 ps (D helix)), purple (3.2 ps (A and C helices)), and ochre (4 ps (EF linker)). (*b*) At the residue level, residues W14, G25, **N26**, L29, I30, D44, R45, F46, S58, **E59**, D60, V66, **V68**, L69, L89, A90, **H93**, A94, **K96, H97**, I107 and **heme** reach their t_c within 1 ps (bold-faced residues are those with $t_c = 400$ fs). (*c*) $t_c > 6$ ps for residues S3, E4, G5, E6, T51, H81, E83, A110, H116, S117, G121, F123, G124, A125, D126, A127, and Q128, among which G5, F123 and G124 culminate at >8 ps.

$\langle \Delta\vec{\mathbf{r}}_a(t) \rangle_f$ recovers back to zero at time infinity through an oscillatory process (Fig. 2 *e*). The oscillation is found to come from the underdamped high-frequency modes (Fig. S6). As we remove the high-frequency contribution from the covariance matrix, $\langle \Delta\dot{\vec{\mathbf{r}}}_j(0)\Delta\vec{\mathbf{r}}_i(t) \rangle_0$, using Eqs. S5-6 and S5-7 in the Supporting Material, the previously observed oscillatory relaxations subside and are smoothened out relative to those in Fig. 2 *e* (Fig. S6).

The fastest disseminators are evolutionarily conserved and essential to the rebinding kinetics

The LDT calculated for every residue i as a function of its distance from the mass center of Mb (d_i) is shown in Fig. 4. The 10 residues with the shortest LDTs, implying fast propagation of mechanical signals throughout the molecule, are ordered by rank as follows (LDT (ps), conservation (indicated by underscoring)): L29 (15.2, 8), A110 (15.2, 4), V68 (15.8, 9), I107 (16.8, 9), I30 (19.0, 4), L72 (19.2, 6), S108 (20.2, 6), P106 (20.6, 7), G73 (20.8, 5), and G109 (21.0, 6). The conservation scores are taken from the ConSurf website (24). Score 9 is the most conserved and 1 is the least. Three of the top four disseminators have a conservation >8. In fact, seven of the top 10 disseminators have a conservation >6, whereas on average only ~50% of Mb residues have a conservation >6. Although it is not

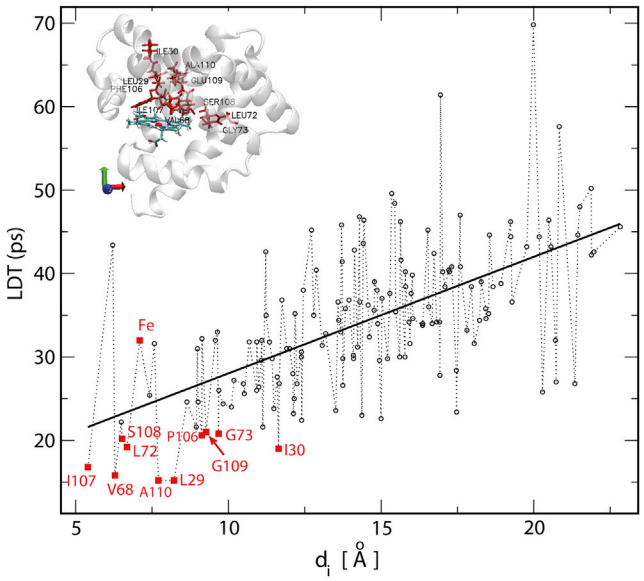

FIGURE 4 LDT of each residue as a function of its distance from the center of mass. Open circles represent the C_α atoms of all residues and solid squares indicate the 10 residues (I107, V68, S108, L72, A110, L29, P106, Q109, G73, and I30) with the shortest LDTs and the ferrous ion in the heme. The solid black line is the linear regression for all the C_α atoms with a correlation coefficient of 0.61. (*Inset*) Relative positions of 10 shortest-LDT residues. The axes indicate the reference coordinates for the impulse- force direction, with x, y, and z axes are colored red, green, and blue, respectively. The direction of impulse force varies as a function of orientational angles (θ, ϕ) in polar coordinates, where θ is the angle between the force direction and the z axis and ϕ is the angle between the projection vector of the force in the xy plane and the x axis. To see this figure in color, go online.

surprising that a general trend toward longer LRT is assumed for residues drifting away from the mass center of this globular protein (Fig. 4), the top disseminator, L29, which is 8.2 Å from the mass center and 9.9 Å from the ferrous ion in the heme, is highly conserved in evolution. L29 was previously shown to have a higher impact on oxygen-binding kinetics than the conserved distal H64 residue that orients the gas molecules. L29F was shown to boost oxygen affinity by 15-fold, whereas H64Q lost affinity by five-fold compared to the wild-type (25). On the other hand, it was shown that the mutations of residue 68 alter size and electric field of the distal primary docking site (so-called state B) in regulating the capture and escape of O_2 and NO (26). Random mutagenesis data reported a number of proximal and distal residues whose mutation would dramatically change the O_2/CO rebinding kinetics (23). These residues include L29, I30 (>12 Å from the mass center), V68, I107, and S108, which happen to be reported herein as disseminators.

The longest relaxation in the system described by LRT corresponds to time-resolved x-ray data

On the other hand, it can be understood that the relaxation time (τ) is more a function of the adopted end-state (targeted) conformation than of the culminating time (t_c), simply because $\Phi(t)_a$ is a direct function of the targeted conformation (see Theories and Methods).

In general, residues that stay at the far end of the molecule have the longest relaxation time constants, τ. They are located in helix D (M55 and K56, with $\tau_{K56} = 81.2$ ps), at the beginnings of helices C (H36) and H (A125 and A127–A130), at the N-terminus (L2–E4), and at the C terminus (K147 and G150–Y151) (Fig. S5). On the other hand, the shortest relaxation times were found for residues in the first half of helix E, residues close to the heme on helix C and the CD loop (P37–E41), a few residues on helix B, including L29 ($\tau = 4.4$ ps), and a few residues on helix A (see Fig. S5). In general, fast relaxations occur on the ligand side of the heme of the binding pocket and near the CO molecule. The relaxation time averaged over all the residues, including the heme, is 16.2 ps. The longest, $\tau_{K56} = 81.2$ ps, is consistent with the time-resolved x-ray data, which suggest that the conformational changes of Mb propagate from the heme throughout the entire protein within 150 ps (27).

Comparisons of td-LRT with other methods in revealing signal propagation pathways

A number of studies have sought to elucidate intramolecular signal propagation pathways (28–36). The propagated physical and nonphysical properties can be potential energy (32), kinetic energy (or heat) (28,30,36), mechanical properties (29,35), information (33,37), or coevolved residues (34). A short review of the methods used in those studies and a comparison with our own method are given in the Supporting Material. Basically, the methods can be broadly divided into two categories, signal pathways characterized in evolutionarily, structurally, and/or dynamically equilibrated systems (32,33,35,37) and those characterized in nonequilibrium systems with introduced perturbations (28–31,36). For equilibrium systems, the pathways identified from highly correlated residues are sometimes hard to understand in clear physical terms, namely, what types of signals are initiated and then propagated. Also, pathways characterized in equilibrated systems may imply that intramolecular communications are independent of how signals/perturbations are introduced into the systems (in terms of size, direction, and frequency), which can be questionable.

Nonequilibrium methods, including ours, provide a good alternative for keeping track of the initiation and dissemination of signals. In particular, our LRT-based method has a few distinct advantages. First, no repetitive MD simulations or NMAs need to be carried out for various types of perturbations touched off at different sites, which is a great computational advantage over other methods. For instance, the LDT results shown in Fig. 4 require that the network be perturbed >4000 times, which can be done in hours. Second, the introduction of perturbations as force impulses in our method bears a closer resemblance to the physical nature of ligand/protein binding than exerting oscillatory forces at fixed frequencies (29) or heating residues locally by coupling with a thermal bath (28,31). Finally, our relaxation-time predictions agree qualitatively and quantitatively with site-specific spectroscopic data from UVRR (3) and time-resolved x-ray (38), without the need of rescaling (7). More distinctions can be found in the Supporting Material.

Despite the very different nature of the aforementioned methodologies and LRT, interesting commonalities in communication behaviors have been jointly observed. For instance, the hit time of Chennubhotla and colleagues characterized asymmetric signal propagations from residues i to j and j to i (37), which we found to be true in Mb: the t_c required to propagate signals from heme to residue K96 takes 2.2 ps, but the reverse propagation takes 4.0 ps due to diffusions of perturbations under different contact topology. Also, Chennubhotla and Bahar observed a faster communication in α-helices than in random coils; in a similar way, we observed that the intrahelix signals propagate more rapidly than the inter-(cross-)helical signals (data not shown). Moreover, we report here a colocalization of kinetically important sites and the residues that most rapidly disseminate signals throughout the molecule, whereas Chennubhotla and Bahar noticed that enzyme catalytic sites are fast signal receivers, moreso than ligand-binding and other nonfunctional residues. It will be interesting to see whether important kinetically hot residues in Mb and enzyme active sites are also good signal receivers (instead of efficient broadcasters) based on our methods as

well as on whether the fast communicators or signal receivers of Chennubhotla are evolutionarily conserved.

Effects of mutation on LRT results

Three mutants that demonstrate various degrees of alteration of CO-rebinding kinetics, D60G, H64P, and L29S, are examined. As shown in the experimental results (Fig. S7) (5), the kinetic changes as represented by deviations of the mutant behavior from the WT rebinding behavior in terms of size and shape of the curves, follow a D60G $<$ H64P $<$ L29S relation. At short time limits, the relation is close to D60G $\approx$ H64P $<$ L29S (Fig. S7). Following the protocol described in Theories and Methods, except that we first mutated both bound and unbound Mb in silico and then performed a long energy minimization to ensure that the average forces of atoms were $<10^{-5}$ kcal/mol/Å, we calculated the conformational relaxation for the three mutants (D60G, H64P, and L29S). Results for D60G show that W14-L69 still has slower relaxation than Y146-I99, with relaxation time constants of 35.7 and 26.3 ps, respectively, where 26.3 ps is about three times longer than the relaxation time for Y146-I99 in the WT (7–8 ps). H64P shows a similar relaxation rate, with commensurate time constants of 11.9 and 15.4 ps for W14-L69 and Y146-I99, respectively, where the 11.9 ps for W14-L69 is around four times faster than the relaxation time observed in the WT (45–59 ps). On the other hand, L29S reverses the trend and shows a threefold faster relaxation for W14-L69 (5.3 ps) than for Y146-I99 (14.7 ps). The 5.3 ps relaxation time is ~10 times faster than the observed result, notably the largest change among the three mutants. We therefore find that time-resolved mechanical changes for these three mutants follow a D60G $<$ H64P $<$ L29S relationship.

Earlier (fast) responses appear to be robust and minimally influenced by the mutants. The t_c of W14 changes only moderately among the three mutants. D60G changes the least, from 2.0 (for WT) to 1.8 ps, whereas for both H64P and L29S, t_c increases by 0.6 ps. On the other hand, t_c of Y146 does not change at all for any of these three mutants, maintaining a fast response of 1.4 ps. Examining the fast responses of the secondary structure elements in these mutants by their t_c values reveals an order of mechanical propagation similar to that found in WT, with the exception of one helix per mutant (see Table S1 for details). However, as we examine the 10 residues in each mutant that bear the shortest LDTs, it is found that 7 of 10 are conserved for D60G and H64P relative to those found in WT, whereas only 5 of 10 are conserved for L29S (Table S2), which follows a D60G $\approx$ H64P $<$ L29S relation. Hence, we do observe appreciable sensitivity of td-LRT to chemical modifications, which is in agreement with site-specific kinetics data for the three mutants we examined.

Slow modes sensible to particular perturbations are selected to drive the overall conformational changes

Under the LRT framework, selected modes respond to specific perturbations to drive conformational change. As shown in Eqs. S9-3 and S9-4 in the Supporting Material, for an intrinsic mode to be highly excited, there must be a joint contribution of two terms, $1/\omega_k^2$ and S_k, where ω_k is the frequency of mode k and the scalar $S_k = \mathbf{u}_{k,j}^T \hat{\mathbf{f}}_j$ measures how much a certain mode is excited by force perturbations. Slow modes by definition carry a large $1/\omega_k^2$ value. S_k is the dot product of two unit vectors, where $\mathbf{u}_k$ points in the directions of mode k and $\hat{\mathbf{f}}$ points in the directions of the perturbation forces. Among the slow modes, when a given mode moves in close accord with the direction of applied forces, this mode is likely to be excited and to contribute significantly to the overall conformational change. This is why mode 11 is excited the most in response to the CO binding relative to the other slow modes (Fig. S4).

Decomposition of the covariance matrix, sampled for unbound proteins, reveals a motional repertoire spanned by corresponding normal modes whereby proteins reconfigure themselves in the absence of ligands, generating a number of preexisting equilibrium states (5,39). The perturbation excites a subset of the modes in the repertoire due to their consistency with the perturbation (high S_k values). Within the LRT, induced fit is understood as a perturbation-selected combination of preexisting motions (normal modes) that are readily sampled in the unperturbed state. When the perturbation is local and vanishingly small, S_k is small in every normal mode, and therefore, $1/\omega_k^2$ takes dominance in the contribution to $\langle \Delta\vec{\mathbf{r}}_i \rangle_f$, hence the robustness of slow modes. This is when the binding scheme retrieves to the notion of preexisting equilibrium (5,39). This was the case in the study of Ikeguchi and co-workers, where the ligand applied forces on a single atom in ferric binding protein (FBP) (2) (a local and small perturbation). Without applying LRT, we examined the intrinsic dynamics of FBP using the elastic network model (2,22) and found that the slowest mode can already explain 73% of the observed conformational changes, with the first two slowest modes explaining 85% and the slowest four explaining 90% of the observed conformational change. When the perturbation is extremely localized, the good agreement between LRT-predicted and observed conformational changes is a manifestation of the intrinsic preference of proteins in motions, indicated by the slowest modes.

Robustness and applicability of the td-LRT

It has been found that the function-promoting, slow normal modes are highly robust against small conformational changes and different modeling/simulation methods (21,40–45). Therefore, in our opinion, the most robust

component in the LRT is the low-frequency normal modes that predominantly comprise the covariance matrix. LRT is generally thought to perform better for proteins with small conformational changes, such as the herein reported Mb system. However, this alone may not fully explain the theoretical reproduction of the UVRR and time-resolved x–ray data. From $\mathbf{f}_j = k_B T \sum_i \langle \Delta\vec{\mathbf{r}}_j \Delta\vec{\mathbf{r}}_i \rangle_0^{-1} \langle \Delta\vec{\mathbf{r}}_i \rangle_f$ (see section titled Time-independent LRT) and Eq. S9-2, one can see that the higher the correlation between the observed conformational changes ($\langle \Delta\vec{\mathbf{r}}_i \rangle_f$) and fast motions (i.e., the eigenvectors associated with high ω_k), the larger are the perturbation forces that can be produced. If many high-frequency motions contribute to the observed conformational changes, the forces would be unnaturally large at room temperature and therefore would invalidate the applicability of LRT. However, our study shows that slow modes contribute dominantly to the observed conformational changes (e.g., mode 11 in Fig. S4), resulting in reasonable forces surrounding the binding pocket. On the other hand, early responses, described by impulse forces, can be even more robust than the slower conformational relaxation, since it is independent of the ending conformation. For instance, our results have shown (in the main text) that the predicted early responses for Trp^{14} and Tyr^{146} are robust against slightly different perturbation forces derived from using different ending structures (CO_min1 or the x-ray-solved bound form of Mb (1A6G)).

LRT was also thought to be unable to describe large conformational changes involving barrier crossing and trespassing energy minima. However, with proper approximations, the linear region could extend beyond what was previously perceived, as supported by both theories and successfully applications (2,7). At the theoretic end, Essiz and Coalson have shown that the assumption of small perturbations in LRT can be dropped as long as the molecule deforms over a harmonic energy landscape (by directly integrating Eq. S2-2 in the Supporting Material with a harmonic Hamiltonian, we have also derived an expression to reach a similar conclusion (data not shown)) (7). Coarse-grained models, bearing a harmonic nature, provide such a lengthened linear region, as demonstrated by Ikeguchi and Kidera, who show that conformational changes (3.1 Å) of the protein citrate synthase (874 residues) can be successfully explained by anisotropic-normal mode-integrated LRT (2). Quasiharmonic analyses (or principal component analysis) on MD trajectories could also provide such a lengthened linear region, again as demonstrated by Ikeguchi and Kidera for FBP and citrate synthase (2), as well as by Essiz and Coalson for FBP and T4-lysozyme (7). However, further studies on larger molecular systems are needed to establish the applicable limits for LRT.

CONCLUSION

The encouraging agreement between spectroscopic (UVRR and x-ray) data and theoretical calculations leads to the belief that the nature of ligand-induced protein primary response is a delicate interplay, described by LRT, between the intrinsic dynamics (capable of generating all possible preexisting equilibrium states) and the perturbation forces (the source of induced fit). The formula also explains at the theoretical level why certain modes are excited during ligand binding. The theories describe the physical nature of mechanical signal propagation within biomolecules, and residues with short LDT seem to be evolutionarily conserved. Overall, the general theories provided here serve as a simple scaffold for enzymologists, medicinal chemists, and protein scientists to understand protein-ligand binding (and protein-protein binding) by reconciling the paradigms of preexisting equilibrium and induced fit.

SUPPORTING MATERIAL

Two tables, seven figures, three movies, Supporting Methods and Discussions are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00795-4.

LWY and AK thank Ministry of Science and Technology of Taiwan for the support (Grant number: MOST 102-2627-M-007-003) and grant-in-aid provided by Japan Society for the Promotion of Science to conduct the research.

We appreciate National Center for High-performance Computing (NCHC) in Taiwan for providing the computational resource for our MD simulations and friction calculations. We also thank Dr Yasuhisa Mizutani for fruitful discussions.

REFERENCES

1. Seno, Y., and N. Gō. 1990. Deoxymyoglobin studied by the conformational normal mode analysis. II. The conformational change upon oxygenation. *J. Mol. Biol.* 216:111–126.
2. Ikeguchi, M., J. Ueno, …, A. Kidera. 2005. Protein structural change upon ligand binding: linear response theory. *Phys. Rev. Lett.* 94: 078102.
3. Sato, A., Y. Gao, …, Y. Mizutani. 2007. Primary protein response after ligand photodissociation in carbonmonoxy myoglobin. *Proc. Natl. Acad. Sci. USA.* 104:9627–9632.
4. Causgrove, T. P., and R. B. Dyer. 1996. Picosecond structural dynamics of myoglobin following photolysis of carbon monoxide. *J. Phys. Chem.* 100:3273–3277.
5. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA.* 102:18908–18913.
6. Manson, A. C., and R. D. Coalson. 2012. Response of rotation-translation blocked proteins using Langevin dynamics on a locally harmonic landscape. *J. Phys. Chem. B.* 116:12142–12158.
7. Essiz, S. G., and R. D. Coalson. 2009. Dynamic linear response theory for conformational relaxation of proteins. *J. Phys. Chem. B.* 113: 10859–10869.
8. Amemiya, T., R. Koike, …, A. Kidera. 2011. Classification and annotation of the relationship between protein structural change and ligand binding. *J. Mol. Biol.* 408:568–584.
9. Amemiya, T., R. Koike, …, M. Ota. 2012. PSCDB: a database for protein structural change upon ligand binding. *Nucleic Acids Res.* 40:D554–D558.
10. McQuarrie, D. A. 2000. Statistical Mechanics. University Science Books, Sausalito, CA.

11. Chandrasekhar, S. 1943. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* 15:1–89.
12. Hayward, S., A. Kitao, …, N. Go. 1993. Effect of solvent on collective motions in globular protein. *J. Mol. Biol.* 234:1207–1217.
13. Brooks, B., and M. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. USA.* 80:6571–6575.
14. Go, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.* 80:3696–3700.
15. Kitao, A., F. Hirata, and N. Gō. 1991. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* 158:447–472.
16. Yang, L. W., E. Eyal, …, A. Kitao. 2009. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics.* 25:606–614.
17. Vojtechovský, J., K. Chu, …, I. Schlichting. 1999. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.* 77:2153–2174.
18. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
19. Kabsch, W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A.* 32:922–923.
20. Eckart, C. 1935. Some studies concerning rotating axes and polyatomic molecules. *Phys. Rev.* 47:552–558.
21. Yang, L. W., E. Eyal, …, I. Bahar. 2007. Insights into equilibrium dynamics of proteins from comparison of NMR and x-ray data with computational predictions. *Structure.* 15:741–749.
22. Yang, L. W., A. J. Rader, …, I. Bahar. 2006. oGNM: online computation of structural dynamics using the Gaussian Network Model. *Nucleic Acids Res.* 34:W24–W31.
23. Huang, X., and S. G. Boxer. 1994. Discovery of new ligand binding pathways in myoglobin by random mutagenesis. *Nat. Struct. Biol.* 1:226–229.
24. Ashkenazy, H., E. Erez, …, N. Ben-Tal. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38:W529–W533.
25. Zhao, X., K. Vyas, …, D. F. Bocian. 1995. A double mutant of sperm whale myoglobin mimics the structure and function of elephant myoglobin. *J. Biol. Chem.* 270:20763–20774.
26. Nienhaus, K., P. Deng, …, G. U. Nienhaus. 2003. Structural dynamics of myoglobin: ligand migration and binding in valine 68 mutants. *J. Biol. Chem.* 278:42532–42544.
27. Hummer, G., F. Schotte, and P. A. Anfinrud. 2004. Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA.* 101:15330–15334.
28. Ota, N., and D. A. Agard. 2005. Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.* 351:345–354.
29. 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.
30. Leitner, D. M. 2009. Frequency-resolved communication maps for proteins and other nanoscale materials. *J. Chem. Phys.* 130:195101.
31. 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. *J. Chem. Phys.* 135:065103.
32. Kong, Y., and M. Karplus. 2009. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins.* 74: 145–154.
33. Chennubhotla, C., and I. Bahar. 2006. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol. Syst. Biol.* 2:36.
34. Lockless, S. W., and R. Ranganathan. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science.* 286:295–299.
35. Ghosh, A., and S. Vishveshwara. 2007. A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl. Acad. Sci. USA.* 104:15711–15716.
36. Yu, X., and D. M. Leitner. 2005. Heat flow in proteins: computation of thermal transport coefficients. *J. Chem. Phys.* 122:54902.
37. Chennubhotla, C., and I. Bahar. 2007. Signal propagation in proteins and relation to equilibrium fluctuations. *PLOS Comput. Biol.* 3:1716–1726.
38. Schotte, F., M. Lim, …, P. A. Anfinrud. 2003. Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science.* 300:1944–1947.
39. Okazaki, K., and S. Takada. 2008. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. USA.* 105:11182–11187.
40. Kitao, A., and N. Gō. 1991. Conformational dynamics of polypeptides and proteins in the dihedral angle space and in the cartesian coordinate space: normal mode analysis of deca-alanine. *J. Comput. Chem.* 12: 359–368.
41. Hayward, S., and N. Go. 1995. Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.* 46:223–250.
42. Ma, J. 2005. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure.* 13: 373–380.
43. Yang, L. W., and I. Bahar. 2005. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure.* 13:893–904.
44. Nicolay, S., and Y. H. Sanejouand. 2006. Functional modes of proteins are among the most robust. *Phys. Rev. Lett.* 96:078104.
45. Bahar, I., T. R. Lezon, …, E. Eyal. 2010. Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* 39: 23–42.