Origin of Protein Quake: Energy Waves Conducted by a Precise Mechanical Machine

Huiyu Li, Shanshan Wu and Ao Ma*

Richard Loan and Hill Department of Biomedical Engineering
The University of Illinois at Chicago
851 South Morgan Street
Chicago, IL 60607

*correspondence should be addressed to:
Ao Ma
Email: aoma@uic.edu
Tel: (312) 996-7225

Abstract

A long-standing challenge in protein biophysics is to understand protein quake in myoglobin—the structural dynamics responsible for redistributing the excess heme energy after photolysis. Despite extensive efforts, the molecular mechanism of this process remains elusive. Using the energy flow theory, we uncovered a fundamental new phenomenon: the heme energy is redistributed by sinusoidal waves with a ubiquitous fundamental frequency and two overtones. The energy waves emanate from the heme into the myoglobin backbone via a conduit of five consecutive dihedrals of the proximal histidine, then travel quickly along the backbone to reach sidechains across the protein. This mechanism is far more effective than the diffusion-based mechanism from previous studies because waves are systematic while diffusion is random. To propagate energy waves, coordinates must cooperate, resulting in collective modes that are singular vectors of the generalized work functional. These modes show task partitioning: a handful of high-energy modes generate large-scale breathing motion, which loosens up the protein matrix to enable hundreds of low-energy vibrational modes for energy transduction.

Proteins are the building blocks of biological systems responsible for most biological functions. Understanding the mechanism of protein function is of paramount importance. The central dogma of protein science is that structure determines function. The critical link between structure and function is the functional dynamics [1, 2] that carries out protein function: structure determines function because the specific structure enables the desired functional dynamics. Understanding protein functions requires understanding the detailed molecular mechanism of functional dynamics.

The physical picture behind functional dynamics is that a protein has a multitude of functional structures, corresponding to basins in the underlying energy landscape that control the behavior of the protein [3]. Transitions between these functional structures, which are carried out by functional dynamics, are often required for protein function. Following the pioneering work by Frauenfelder and co-workers [2], this idea has played a central role in experimental and computational studies of protein biophysics for the past few decades [4-8].

One of the first and most studied functional dynamics is the protein quake, a concept derived from pioneering studies of the structural dynamics of myoglobin (**Mb**) after photolysis by Ansari et al [1]. The physical picture of protein quake is that the deposition of excess energy at a local site causes a strain in a protein, which then propagates through the protein matrix like an earthquake propagating through the mantle of earth. There are two essential hypotheses in this immensely appealing picture: 1) wave-like energy transduction through the protein matrix, 2) quake-like structural dynamics. Both ideas have attracted long-lasting and recurring interests and efforts from experimentalists and theorists alike.

The energy transduction during protein quake has been the subject of many spectroscopic studies since the pioneering work by Hochstrasser and co-workers in 1978 [9-16]. These experiments were complemented by computational studies dating back to the work by Henry et al in 1986 [17-19]. In a series of classical studies, Straub and co-workers investigated mechanisms of both direct transfer of excess heme energy to the solvents through the solvent-exposed isopropionate sidechains and energy transfer within the protein [19-22]. The latter employs three mechanisms: 1) “through projectile” due to interactions between the dissociated ligand and protein sidechains, 2) “through space” due to interactions between the heme and sidechains in the heme pocket, and 3) “through

bond", with the bond between the heme iron and the proximal histidine (**His93**), the only bonded connection between heme and the protein, as the starting point. The general mechanism emerged, which combines all these factors together, is an anisotropic diffusion process that completes in ~5 ps [19, 22]. On the other hand, as originally pointed out by Lian et al [16], the slow time scale of diffusion contradicts the observed fast energy transfer, leading to the hypothesis of ballistic energy transfer by collective modes. This hypothesis was supported by an intriguing computational study of a related process in lactate dehydrogenase by Schwartz and co-workers [23]. They found that the transduction of excess thermal energy from the active-site proceeds preferentially along the directions of the promoting vibrations that are required for catalysis of hydride transfer—the main function of this enzyme. The existence of preferred axis of thermal energy transfer challenges the diffusion-centric picture of energy transduction in proteins. Moreover, it suggests a strong link between thermal energy transfer and enzymatic function. The open question is: What are the collective modes responsible for energy transfer in proteins?

On the structural dynamics side, recent advancements in X-ray free electron lasers (**X-FEL**) made it possible to "film" molecular movies of ultrafast protein dynamics [24], kindling the efforts to capture the protein quake in action [25]. Time-resolved femtosecond crystallography, complemented by QM/MM simulations, showed collective response of the protein within 500 fs of photolysis of Mb that originates from the coupling of heme vibrations to global protein modes [26], but the specific nature of the protein modes is unclear. Similarly, time-resolved solution phase small- (**SAXS**) and wide-angle X-ray scattering (**WAXS**) showed an increase of the radius of gyration (R_g) of Mb within 1 ps, followed by damped oscillations with a period of ~3.6 ps [27]. Another study using time-resolved SAXS and WAXS detected a single pressure peak propagating through the protein within 500 fs [28].

Despite many years of efforts and the abundance of information available on both aspects of protein quake, some critical links are still missing: 1) Is energy transduction wave-like as originally conjectured or diffusion-like as suggested by computational studies? 2) What is the molecular pathway for the energy transduction and what is the role of each residue or coordinate on this pathway? 3) What are the collective modes responsible for the energy transduction during a protein quake? These are the questions we strive to answer in the current paper.

We focused on the prototype of protein quake: the energy transduction and structural dynamics of Mb after photolysis. This process has two parts. After the heme absorbs a photon, a fraction of the photon energy is used to sever the Fe-CO bond on a femtosecond time scale, and the remaining photon energy is redistributed among the vibrational modes of the heme almost instantly on a femtosecond time scale [19, 29]. As a result, the heme has a significant amount of excess energy, which is estimated to be 88 kcal/mol = 368 kJ/mol by Sagnella et al and 81 kcal/mol by Henry, Eaton and Hochstrasser in their first MD simulation of the cooling of heme [17, 19]. This excess heme energy is then redistributed to the rest of the protein matrix, which essentially completes in 5 ps [10, 12, 15, 19]. This heme cooling process is the main part of the protein quake and the subject of the current study.

Most computational studies assumed that heme cooling is dominated by diffusion and focused on delineating the microscopic models for the friction kernel, which is the determining factor for diffusion [19, 30-32]. In contrast, we used energy flow theory to analyze the energy changes in all the coordinates in Mb and the generalized work functional method to identify cooperativity between individual coordinates. While previous computational studies of heme cooling were carried out in Cartesian coordinates, all the analyses in the current work were conducted in internal coordinates. This is because, from our studies of activated processes in peptides, we found that the energy flows of internal coordinates can reveal valuable mechanistic insights [33-36]. For instance, the two major reaction coordinates of the $C_{7eq} \rightarrow C_{7ax}$ isomerization of an alanine dipeptide, the coordinates that control the isomerization dynamics, have high energy flows, while the non-reaction coordinates have essentially zero energy flows [33, 34]. The leading singular vector of the generalized work functional tensor of this process, computed in internal coordinates, is the rigorously defined one-dimensional reaction coordinate [35]. In contrast, the energy flows of Cartesian coordinates during this process are homogeneous and essentially noise. The reason for the difference between internal and Cartesian coordinates is that the former are the natural coordinates for describing protein motions because they automatically satisfy all the constraints from bonded interactions. Their energy flows are the energy cost of the motion of individual coordinates and reflect their importance. In contrast, motions of Cartesian coordinates are dominated by strong restraint forces

due to bonded interactions and any mechanistically relevant information is buried by the strong noised from the restraint forces.

Contrary to the diffusion-centric mechanism from previous studies [17, 19], we found that the excess heme energy is redistributed by energy waves with a ubiquitous fundamental frequency and two overtones. The excess heme energy is channeled into the Mb backbone by the proximal histidine (**His93**) and then dispatched across the protein matrix. To propagate the energy waves, different coordinates must cooperate closely and Mb acts like a precise mechanical machine. The mechanism of this machine is best described by the singular vectors of the generalized work functional. Contrary to diffusion, waves are directional and systematic, thus they are more effective in distributing the heme energy. While diffusion-centered mechanism suggests that redistribution of heme energy resembles ink spreading in still water, the mechanism we found alludes to tiny objects ridding on ripples. The latter travels much faster than the former.

Results

To simulate heme cooling in Mb after photolysis, extra kinetic energy is deposited to heme atoms $t = 0$ to mimic the excess energy from absorbing a photon. Afterwards, the system is simulated using constant energy molecular dynamics simulation (details in Methods).

Energy flow theory. Our analysis is based on the potential energy flow (**PEF**) of individual coordinates (details in Supplementary Information (**SI**) and refs. [33, 34]). The PEF of a coordinate q_i along a dynamic trajectory α is the mechanical work done on q_i [33]:

$$\Delta W_i(t_1 \to t_2; \alpha) = \int_{q_i(t_1;\alpha)}^{q_i(t_2;\alpha)} F_i dq_i = -\int_{q_i(t_1;\alpha)}^{q_i(t_2;\alpha)} \frac{\partial U(\vec{q})}{\partial q_i} dq_i \quad (1),$$

where $U(\vec{q})$ is the potential energy of the system, $\vec{q}$ is the position vector of the system in the configuration space, and F_i is the total force, including forces from both solvents and the protein, exerted on q_i. Because $\Delta W_i(t_1 \to t_2; \alpha)$ is the change in $U(\vec{q})$ caused by the motion of q_i alone, it is the energy cost of the motion of q_i. A change in $U(\vec{q})$ can be exactly decomposed as: $\Delta U(t_1 \to t_2; \alpha) = -\sum_i \Delta W_i(t_1 \to t_2; \alpha)$, where the summation is over all the coordinates. This ensures that there is no missed count or overcount of any factor in the PEF of each coordinate.

To understand the mechanism of a protein quake, we compute PEFs averaged over an ensemble of trajectories, each trajectory simulates an instance of protein quake:

$$\langle \Delta W_i(0 \rightarrow t) \rangle = \frac{1}{N} \sum_{\alpha=1}^{N} \Delta W_i(0 \rightarrow t; \alpha) \quad (2),$$

where $t = 0$ is set to be the time when the excess kinetic energy is deposited to the heme, N is the number of trajectories in the ensemble.

Energy transduction by waves. The PEFs of an internal coordinate along different trajectories show distinct patterns (Fig. S1). Consequently, for each coordinate q_i, we used k-means clustering (details in SI) to divide 8,000 trajectories into six ensembles $E_s(q_i); s \in \{0, \dots, 5\}$. The PEFs along trajectories in each ensemble resemble one pattern. For each coordinate q_i, $E_0(q_i) \simeq E_1(q_i) \simeq 3{,}300$, $E_2(q_i) \simeq E_3(q_i) \simeq 540$, $E_4(q_i) \simeq E_5(q_i) \simeq 160$. For two different coordinates q_i and q_k, $E_s(q_i) \neq E_s(q_k)$. Moreover, $E_s(q_i) \cap E_r(q_k) \neq 0; r, s \in \{0, \dots, 5\}; r \neq s$; its value indicates how strongly the motions of q_i and q_k are correlated.

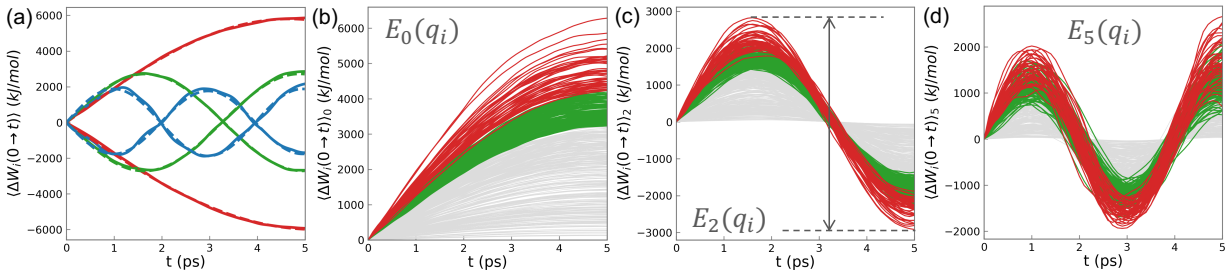

Figure 1: PEFs of backbone dihedrals are sine waves. (**a**) PEF of a backbone dihedral averaged over all the six ensembles. Dashed lines are the best fits to $A_i \sin(\omega_i t + \phi_i)$. (**b** to **d**) PEFs of all the backbone dihedrals averaged over $E_0(q_i), E_2(q_i)$ and $E_5(q_i)$, respectively.

Figure 1a shows the average PEFs of a backbone dihedral for all its 6 ensembles. Strikingly, the PEFs are wave-like. The dashed lines are the best fit of each PEF to the functional form $A \sin(\omega t + n\pi); n = 0, 1$. The fits are nearly perfect, confirming that the PEFs are sine waves, with the 6 ensembles corresponding to 0.25, 0.75 and 1.25 period and initial phase of 0 and π, respectively. The fitted parameters (Table S1) correspond to a fundamental frequency of $0.32 \pm 0.01\ ps^{-1}$ and its third and fifth overtones.

Figure 1 shows the average PEFs for all the backbone dihedrals for $E_0(q_i), E_2(q_i), E_4(q_i)$. All the PEFs are sine waves with essentially the same fundamental frequency. Each internal coordinate has 6 different excitation modes, corresponding to different combinations of overtones (1, 3 or 5)

and phases (0 or π). In a single trajectory, different coordinates adopt different excitation modes; the initial configuration and momenta of Mb determines which excitation mode each coordinate adopts. These results support the idea that a protein quake uses waves for energy transduction, as Ansari et al originally conjectured in 1985 [1]!

We note that the PEFs of many backbone dihedrals and bond angles have a magnitude of thousands of kJ/mol. This is because a small change in a backbone internal coordinate can potentially lead to large-scale rigid motion of the Mb. For example, a change in a backbone dihedral, with all the other coordinates in the system kept fix, leads to a relative motion between the two parts of the protein on the opposite sides of the bond of rotation that defines the dihedral. This kind of motion will involve significant change in system potential energy, resulting in the high PEFs of backbone dihedrals.

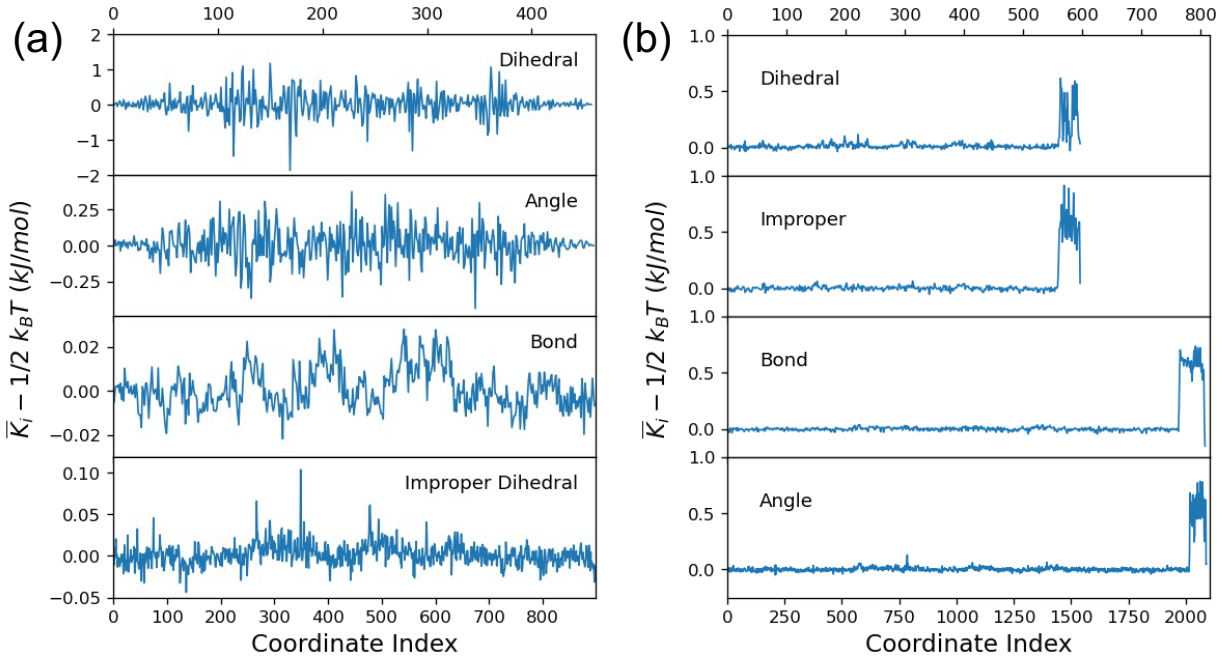

Figure 2: Deviations of kinetic virials of different coordinates from the equilibrium value. **(a)** Kinetic virial of backbone coordinates. X-axis marks at the top are indices of dihedrals, bond angles and bonds; x-axis marks at the bottom are indices of improper dihedrals. **(b)** Kinetic virials of sidechain coordinates. All the coordinates with high kinetic virials belong to the heme. X-axis marks at the top are indices of dihedrals, improper dihedrals and bonds. X-axis marks at the bottom are indices of bond angles.

Energy level of individual coordinates. Each coordinate appears to have an intrinsic capacity for PEF that is independent of its mode of excitation. For a coordinate q_i, the maximum of $\langle \Delta W_i(0 \to t) \rangle_0$ is much higher than the maximum of $\langle \Delta W_i(0 \to t) \rangle_2$, but $\langle \Delta W_i(0 \to 5\ ps) \rangle_0 =$

$\langle \Delta W_i(1.5\ ps \rightarrow 5\ ps)\rangle_2 = -\langle \Delta W_i(0 \rightarrow 5\ ps)\rangle_1 = -\langle \Delta W_i(1.5\ ps \rightarrow 5\ ps)\rangle_3$ (gray arrow in Fig. 1c). This is true for all the coordinates. This intrinsic PEF capacity of a coordinate can be considered its energy level.

Energy Waves are Driven by Excess Heme Energy. The wave form of the PEFs (i.e. $\langle \Delta W_i(t)\rangle$) indicates that, during heme cooling, Mb undergoes systematic dynamics that are distinct from the random motions of equilibrium fluctuations. The purpose of this systematic dynamics is to disperse the excess heme energy. Unlike a large-scale conformational change that can be easily seen from the protein structure, this systematic dynamics is subtle and only revealed by the PEFs of internal coordinates, which are the potential energy required to sustain the systematic dynamics of individual internal coordinates but not the excess heme energy passing through them. The latter accounts for only a tiny fraction of $\langle \Delta W_i(t)\rangle$.

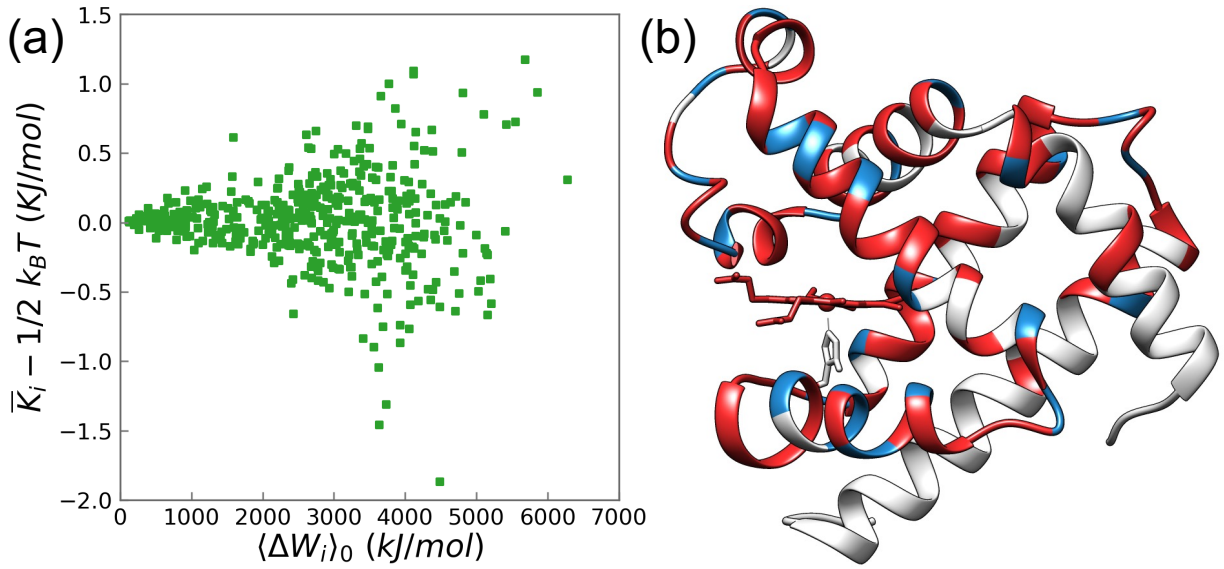

Figure 3: Kinetic virial and PEF. **(a)** Scattered plot showing the correlation between the kinetic virial and PEF of individual backbone dihedrals. **(b)** Red: atoms of backbone dihedrals with $\overline{K}_i - \frac{1}{2}k_BT > 0.25$ kJ/mol; Blue: atoms of backbone dihedrals with $\overline{K}_i - \frac{1}{2}k_BT < -0.25$ kJ/mol.

To find out the excess heme energy passing through a coordinate q_i, we calculated its kinetic virial $K_i = \frac{1}{2}p_i\dot{q}_i$, where $\dot{q}_i$ and $p_i = \partial H/\partial \dot{q}_i$ are the velocity and momentum of q_i, respectively [33]. With this definition, the total system kinetic energy can be decomposed as: $K = \sum_i K_i$. For a Cartesian coordinate x_α, $K_\alpha = \frac{1}{2}p_\alpha \dot{x}_\alpha = \frac{1}{2}m_\alpha \dot{x}_\alpha^2$, is the kinetic energy of x_α. For an internal coordinate q_i, K_i is not the kinetic energy of q_i because $p_i = \sum_j \mathbb{S}_{ij}\dot{q}_j$ is a function of all the

coordinates and velocities in the system. Here, $\mathbb{S}_{ij} = \sum_\alpha m_\alpha \frac{\partial x_\alpha}{\partial q_i} \frac{\partial x_\alpha}{\partial q_j}$ is the structural coupling tensor. We note that K_i can have negative values, which indicates that p_i and $\dot{q}_i$ anti-correlate with each other but does not mean that q_i has very low kinetic energy. To consider the contribution to the total system kinetic energy from q_i, only the magnitude of K_i matters--a larger magnitude means higher importance.

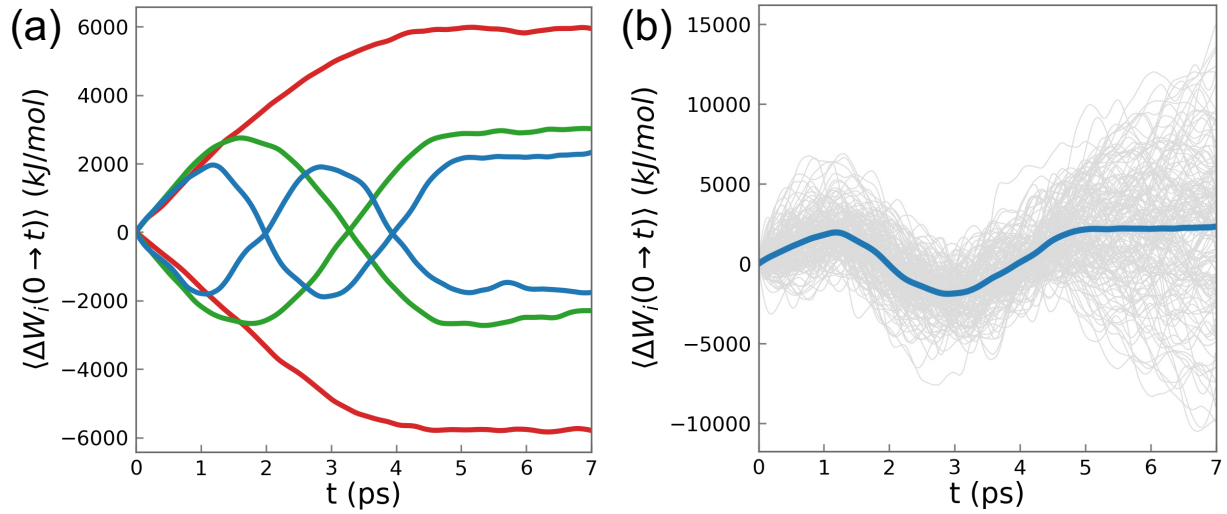

Figure 4: Dephasing of energy waves. **(a)** The PEF of a backbone dihedral (the same coordinate as in Fig. 1a) averaged over all the six ensembles **(b)** The PEFs of a coordinate along individual trajectories in $E_5(q_i)$. The blue line is the average PEF. Note that the PEF curves start to spread out after 5 ps.

A critical feature of K_i is that it satisfies the equipartition relation $\langle K_i \rangle = \frac{1}{2} k_B T$ [33]. Thus, the deviation of $\overline{K}_i = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{T_0} \int_0^{T_0} K_i(k;t) dt$, the average value of K_i during heme cooling, from $\frac{1}{2} k_B T$ indicates the amount of excess heme energy passing through q_i. Here, $N = 8{,}000$ is the number of trajectories, $T_0 = 5$ ps is the length of each trajectory, $K_i(k;t)$ is the kinetic virial of q_i at time t in the k-th trajectory, and $\frac{1}{2} k_B T = 1.25$ kJ/mol at T=300K. Figure 2 shows $\overline{K}_i$ of different types of internal coordinates. Only kinetic virials of heme coordinates and the backbone dihedrals and bond angles show significant deviation from $\frac{1}{2} k_B T$. The excess kinetic virials of heme coordinates are due to the excess heme energy. The excess kinetic virials of backbone dihedrals and bond angles indicate that the backbone is the highway for energy transduction in proteins. The excess heme energy travels quickly along the backbone to be dispatched to the sidechains. Sidechain coordinates do not have excess kinetic virials because they are at the terminal end of

energy transduction; each sidechain coordinate receives only a small fraction of the 368 kJ/mol of excess energy, as there are 7652 internal coordinates in Mb.

Figure 3 shows that, for backbone dihedrals, the deviation of $\overline{K}_i$ from its equilibrium value (i.e. $\overline{K}_i - \frac{1}{2}k_BT$) generally increases with the magnitude of $\langle \Delta W_i \rangle_0$. This suggests that coordinates with high $\langle \Delta W_i \rangle$ play more important roles in dispersing the excess heme energy. Because $\langle \Delta W_i \rangle$ is the energy cost for sustaining the systematic motion of q_i, this shows that the energy waves of the internal coordinates are driven by and used for dispersion of the excess heme energy.

This assumption is supported by the dephasing behavior of the energy waves. Figure 4 shows the energy waves of a backbone dihedral start to lose their coherence after 5 ps, which is the time when the excess heme energy is diminished. Figure 4b shows how the dephasing happen in individual MD trajectories in the ensemble $E_4(q_i)$. The $\Delta W_i(t)$ along a trajectory starts to lose its wave form and spread out after 5 ps. This result shows that the coherence of the energy wave is indeed driven the excess heme energy and the former fades away when the latter diminishes. These results show that we can map out the pathway for the transduction of the excess heme energy as the internal coordinates with high PEFs.

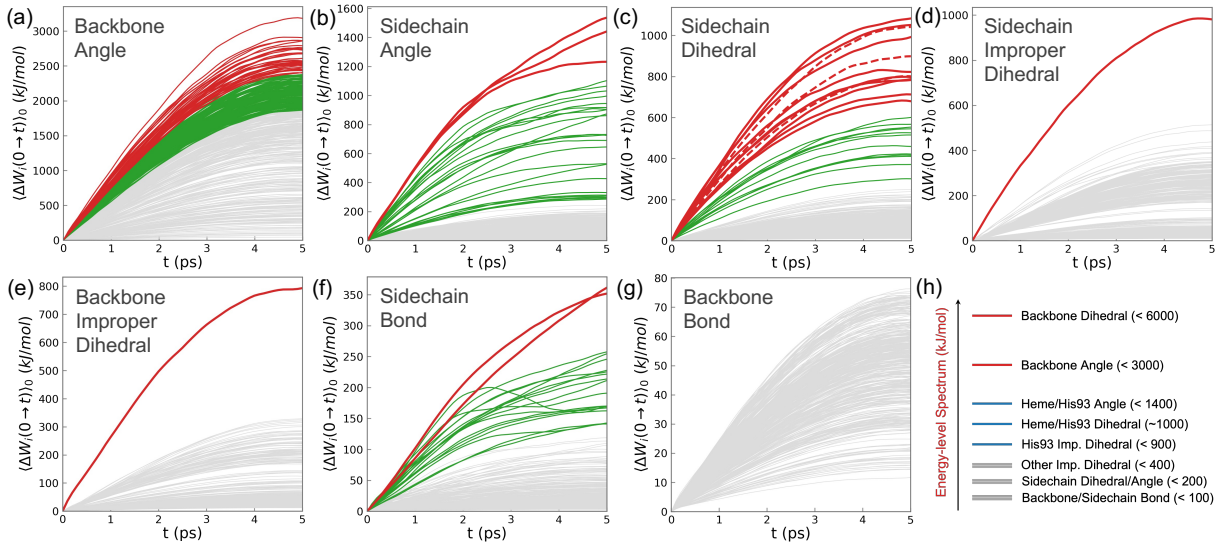

Figure 5: PEFs of different types of internal coordinates averaged over $E_0(q_i)$. (**a** to **g**) PEFs of different types of internal coordinates, colored based on their magnitudes. In (**b**, **c**, **f**), all the red and green curves are PEFs of bond angles, dihedrals and bonds of heme or His93.. In (**d**, **e**), the two red curves are PEFs of improper dihedrals δ_1, δ_2 of His93 shown in Fig. 3d. (**h**) A schematic showing the hierarchy of PEFs of different types of internal coordinates.

Molecular pathway and mechanical mechanism of energy transduction in Mb. Figure 2 shows the PEFs of all the other types of internal coordinates averaged over their own $E_0(q_i)$. There are

two essential features. First, PEFs of different types of coordinates differ significantly and there is a clear hierarchy. The first tier consists of dihedrals and bond angles of the backbone; their maximum PEFs can reach several thousand kJ/mol. The second tier consists of dihedrals, bond angles and improper dihedrals of the heme and His93. Their maximum PEFs can reach 1400 kJ/mol. The third tier consists of the backbone improper dihedrals, and dihedrals, bond angles, improper dihedrals and bonds of sidechains. Their maximum PEFs are well below 300 kJ/mol. In particular, the PEFs of backbone and sidechain bonds are extremely low—below 80 kJ/mol. Clearly, different types of internal coordinates serve distinct functional roles.

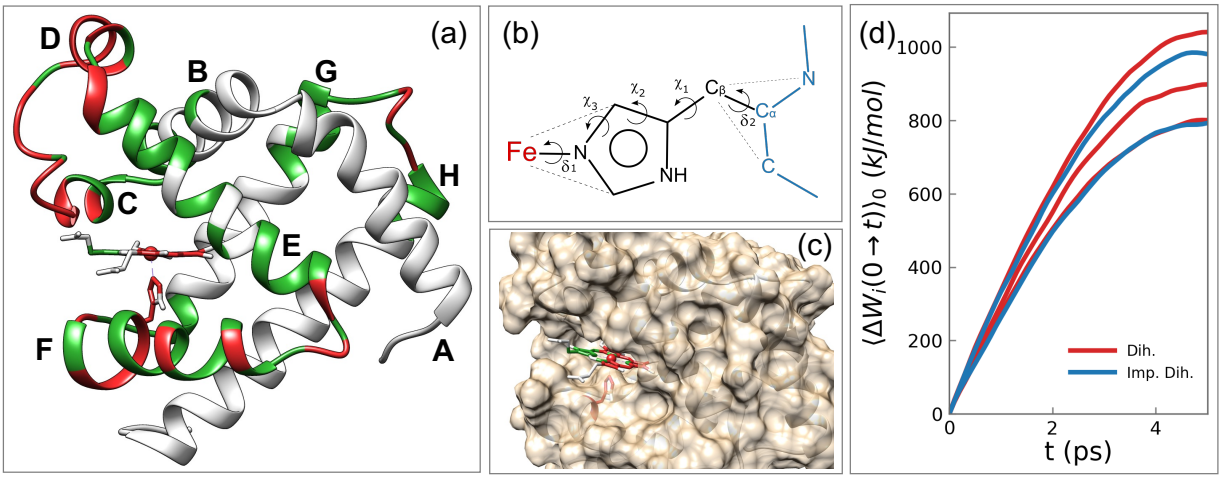

Figure 6: Molecular pathway for energy transduction in Mb. **(a)** Structure of Mb with different coordinates colored based on the magnitudes of their PEFs. **(b)** His93 as the connector between the heme and the protein backbone (atoms in **blue** color). The five internal coordinates that are critical for channeling the heme energy are labeled. **(c)** A close look at the difference in PEFs through the parts of the heme facing the interior and exterior of Mb. **(d)** PEFs through δ_1, δ_2 (**blue**) and χ_1, χ_2, χ_3 (**red**).

Second, the main channel for redistributing the excess heme energy is through His93 into the protein backbone. Here, heme is the energy source, protein backbone is the repository, and His93 is the connector in between. This connector consists of two junctions, one with the heme and the other with the protein backbone, and a duct in between. The duct consists of three dihedrals χ_1, χ_2, χ_3, and the two junctions are improper dihedrals δ_1 and δ_2 (Fig. 3). Both a δ_1 and δ_2 have PEF >800 kJ/mol (Fig. 2d,e), far above the maximum PEFs of all the other backbone and sidechain improper dihedrals, which are below 300 kJ/mol. Basically, the excess heme energy is conducted into His93 via the vibrations of δ_2, then propagates through His93 sidechain via the puckering of the imidazole ring, controlled by χ_3, χ_2 and χ_1, before passing into the backbone via the vibrations of δ_1. The two junctions utilize improper dihedrals instead of other types of internal coordinates because their vibrations require little overall motion of the heme and His93 sidechain, which are

difficult in the crowded environment of the heme pocket. The duct utilizes dihedrals because large-amplitude motion of the imidazole ring is required to convey high energy.

To visualize the molecular pathway for energy transduction, we divide the internal coordinates into groups based on the magnitudes of their PEFs and colored them accordingly (Fig. 2). The groupings are based on comparing PEFs of coordinates of the same type (e.g. we compare PEFs of backbone dihedrals with each other, but do not compare PEFs of dihedrals to PEFs of bond angles). In the Mb structure in Fig. 3a, all the atoms of an internal coordinate are in the same color as its PEF. Therefore, the colors in Fig. 3a show the magnitudes of PEFs through different parts of Mb. The part of heme facing the interior of the heme pocket is "red hot" while the part facing the solvents has much lower PEFs (Fig. 3c). This suggests that heme-protein interactions are much more important for dispensing the heme energy than heme-water interactions, a conclusion different from previous computational studies [19]. The entire His93 side chain is red because it channels the heme energy into helix F, which then propagates along two pathways. The major pathway heads towards the turn between helices F and E and then moves up helix E towards the CD-turn, the part of Mb that experiences the highest PEFs. The minor pathway is towards the turn between helices F and G. Overall, loops connecting helices experience the highest PEFs, while long helices have low PEFs, with helices E and F experiencing higher PEFs than the other helices.

Together, these results suggest that the redistribution of the excess heme energy is carried out by a mechanical machine that employs a hierarchy of energy flow channels. This machine operates similarly to a railroad system: the heme is the central terminal, protein backbone is the railroad, sidechains are the train stations, energy waves are the trains and bits of the excess heme energy are the passengers. After the heme energy is conducted into the backbone through His93, it rides on the energy waves to travel quickly along the backbone. When it reaches a sidechain on the way, a fraction of it is dispatched into the sidechain, with the size of the fraction determined by the size of the sidechain. In this way, heme energy is quickly redistributed over the entire protein matrix. This mechanism is much more effective than diffusion because wave is a directional motion while diffusion is random.

Energy waves require a structural dynamics with global cooperativity. Figure 4a shows the PEFs of all internal coordinates averaged over $E_0(\phi - 124)$, where $\phi - 124$ is the ϕ dihedral of Gly124. Aside from the PEF of $\phi - 124$ itself, PEFs of all but four dihedrals are vanishingly small. This is because $E_0(\phi - 124) \cap E_s(q_i) \simeq E_0(\phi - 124) \cap E_{s+1}(q_i); s = 0,2,4$, leading to strong cancelations and vanishing PEFs. In contrast, for the four dihedrals ψ, ω - 124, ψ of Phe123, and ω of Ala125 (blue lines), $E_0(\phi - 124) \cap E_1(q_i) \gg E_0(\phi - 124) \cap E_s(q_i)$ $(s \in \{0, 2, 3, 4, 5\})$. Therefore, their PEFs averaged over $E_0(\phi - 124)$ are dominated by their corresponding $\langle \Delta W_i(0 \to t) \rangle_1$, leading to the high PEFs in Fig. 4a. This situation indicates strong

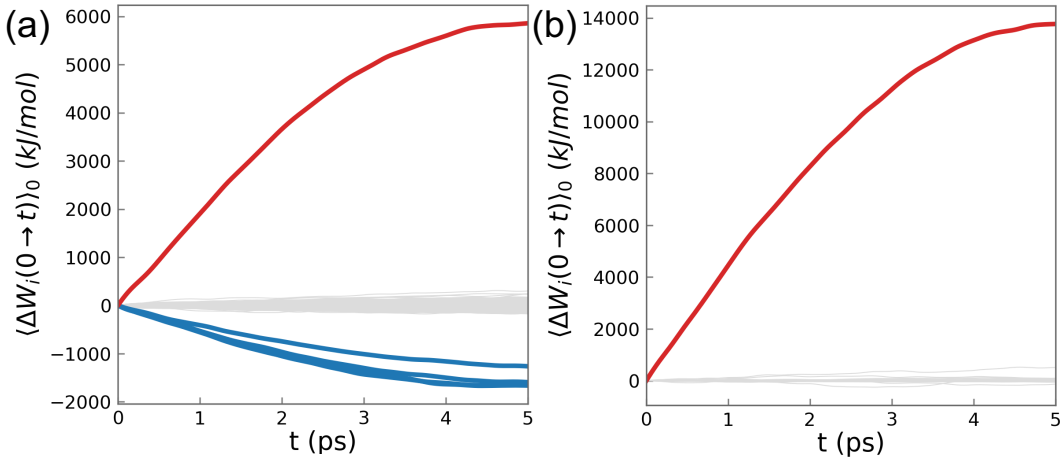

Figure 7: Correlations between different coordinates. **(a)** PEFs of all the internal coordinates averaged over $E_0(\phi - 124)$. **Red**: PEF of $\phi - 124$; **Blue**: PEFs of $\omega, \psi - 124, \psi - 123, \omega - 125$; **Gray**: PEFs of all the other dihedrals. **(b)** PEFs of all the singular coordinates averaged over $E_0(u_0)$. **Red**: PEF of u_0; **Gray**: PEFs of all the other singular coordinates.

correlations between the modes of excitation of $\phi - 124$ and its four neighboring dihedrals. This type of correlation exists extensively between dihedrals of a residue i and dihedrals of residues $i \pm 1$. These local correlations concatenate into a chain of correlations that extends through the entire protein backbone, leading to global cooperativity between internal coordinates.

This global correlation is critical for function. While only 368 kJ/mol excess heme energy needs to be redistributed across Mb, PEFs of individual coordinates (Figs. 1, 2) far exceed this value because they are required for the systematic structural dynamics that is necessary for sustaining and propagating energy waves. The global correlation ensures that the high negative PEFs of one set of coordinates are balanced out by the high positive PEFs of another set of coordinates, keeping

the overall energy cost of the systematic structural dynamics low. This enables redistributing the heme energy via energy waves.

Generalized work functional. The existence of global correlation means the structural dynamics during a protein quake is best described by collective modes. To uncover these modes, we use the generalized work functional [35]:

$$\mathbb{W}_{ij}(0 \to t) = \int_{q_j(0;\alpha)}^{q_j(t;\alpha)} \langle F_i dq_j \rangle = \frac{1}{N} \sum_{\alpha=1}^{N} \int_{q_j(0;\alpha)}^{q_j(t;\alpha)} F_i dq_j \quad (3),$$

which summarizes the impact of F_i on the motion of q_j. The collection of all the $\mathbb{W}_{ij}(0 \to t)$ is an asymmetric tensorial functional $\mathbb{W}(0 \to t) = \int_0^t \langle \vec{F} \otimes d\vec{q} \rangle$ in the configuration space, as $\mathbb{W}_{ij} \neq \mathbb{W}_{ji}$. It summarizes the accumulated effects of the mechanical couplings between different coordinates as a function of the time since photolysis.

Singular coordinates encapsulate the cooperativity. We use singular value decomposition to extract the left singular vectors of $\mathbb{W}(0 \to t)$, which are the directions of forces that have the highest impacts on the motion of the system. Each singular vector u_i is a collective mode, which we refer to as a singular coordinate. Because $\mathbb{W}(0 \to t)$ depends on time, the analytical expression of u_i is time dependent in principle. In practice, we found that expression of u_i is time independent within numerical errors.

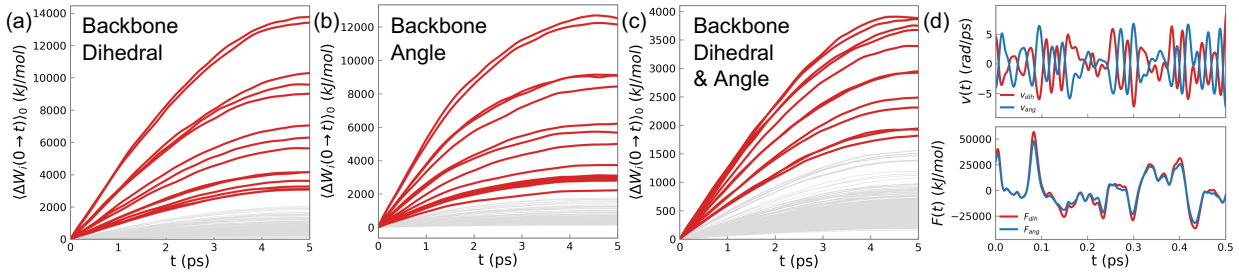

Figure 8: PEFs of singular coordinates. PEFs of all the singular coordinates constructed using all the internal coordinates of heme and His93 and backbone dihedrals **(a)**, backbone bond angles **(b)** and both **(c)**, averaged over $E_0(u_i)$. **(d)** The dihedral (**red**) and bond angle (**blue**) components of the velocity (**upper**) and the force of u_0 (**lower**) along a sample trajectory.

Figure 4b shows the PEFs of all the singular coordinates, constructed from the backbone dihedrals and all the coordinates of heme and His93, averaged over $E_0(u_0)$. Unlike the case in Fig. 4a, only $\langle \Delta W(u_0) \rangle$ has high magnitude, while $\langle \Delta W(u_i) \rangle$ for all $i \neq 0$. This means that the singular coordinates are uncorrelated, thus they serve functional purposes in independent manner. This is

because the correlations between different internal coordinates are converted into the cooperativity between different components of u_i, defined by the coefficients in the analytical expression of u_i.

Figures 5 shows the PEFs of singular coordinates constructed from backbone dihedrals (Fig. 5a), bond angles (Fig. 5b), and the two together (Fig. 5c), aside from the coordinates of heme and His93. While the highest PEFs in the first two cases are far above, the highest PEFs of the third case are on par with the highest PEFs of backbone internal coordinates. The reason is that the PEFs of the bond angles compensate the PEFs of the dihedrals along every single trajectory, lowering the PEF of the singular coordinate. Figure 5d shows the dihedral and bond angle components of the velocity ($\dot{u}_0$) and force (F_0) of $u_0 = \sum_{i=1}^{n_d} c_i h_i + \sum_{j=1}^{n_a} c_j a_j = u_{0,d} + u_{0,a}$ along a trajectory, where n_d, n_a are the total number of backbone dihedrals and bond angles respectively, h_i, a_j are individual dihedrals and bond angles, c_i, c_j are the corresponding coefficients. Accordingly, $\dot{u}_0 = \sum_{i=1}^{n_d} c_i \dot{h}_i + \sum_{j=1}^{n_a} c_j \dot{a}_j = \dot{u}_{0,d} + \dot{u}_{0,a}$, and $F_0 = -\frac{\partial U}{\partial u_0} = -\sum_{i=1}^{n_d} c_i \frac{\partial U}{\partial h_i} - \sum_{j=1}^{n_a} c_j \frac{\partial U}{\partial a_j} = F_{0,d} + F_{0,a}$. While $F_{0,d} \simeq F_{0,a}$ along each trajectory (lower panel in Fig. 5d), $\dot{u}_{0,d}$ and $\dot{u}_{0,a}$ (upper panel) are always of opposite sign and similar magnitudes. Consequently, $dW_{0,d} = F_{0,d}\dot{u}_{0,d}dt$ and $dW_{0,a} = F_{0,a}\dot{u}_{0,a}dt$ are of opposite sign and cancel with each other, significantly lowering $dW_0 = F_0\dot{u}_0 dt$. This precise energy compensation does not exist between individual dihedrals and bond angles; it only exists between the dihedral and bong angle components of the same singular coordinate (i.e. $u_{i,d}$ and $u_{i,a}$). This shows that singular coordinates precisely encapsulate the cooperativity between different coordinates, which is critical for the functional purpose of the protein quake. Therefore, singular coordinates can properly describe the structural dynamics during a protein quake.

Energy-level splitting. The PEFs of singular coordinates (Fig. 5) clearly separate into two groups. A small number of singular coordinates (13) have high PEFs and their "energy levels" are well separated. In contrast, all the other singular coordinates have low PEFs, which are densely packed into a narrow range. This clear difference in PEFs points to different functional roles. We refer to the first group (red lines in Fig. 5) as the functional modes and the second group (gray lines) as the dissipation modes. The clear separation in the "energy levels" upon transformation from internal

to singular coordinates resembles the energy level splitting upon transformation from atomic to molecular orbitals.

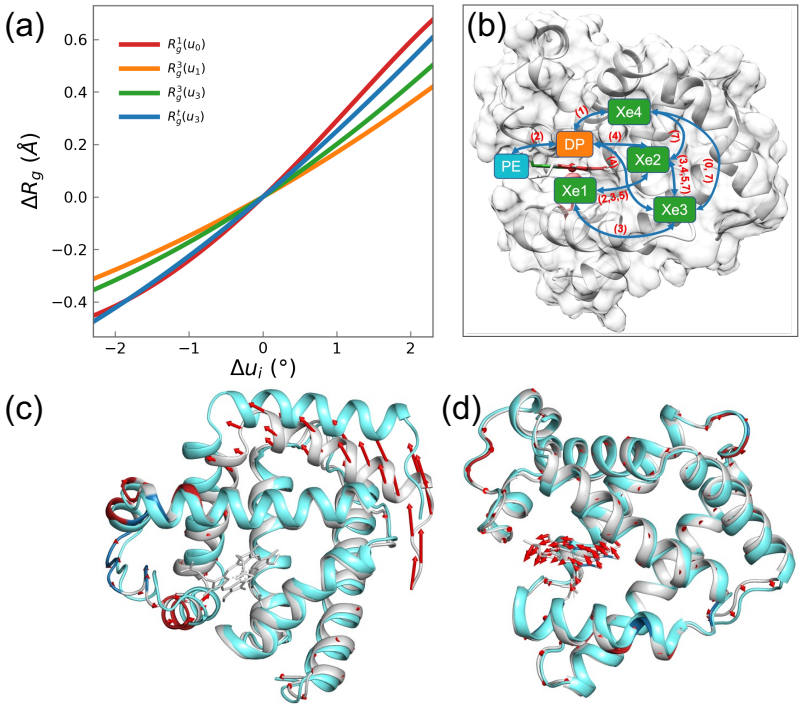

Figure 9: Effects of functional modes. **(a)** Change in R_g caused by displacements along a few functional modes. $R_g^1(u_0)$ means change in the radius of gyration around the first principal axis of Mb caused by displacement along u_0. $R_g^t(u_3)$ means change in the overall R_g. **(b)** A schematic that summarizes the effects of displacements along different functional modes on the connectivity between internal cavities of Mb based on data in Table S2. A **blue** arrow connecting two cavities represents the channel between them. The numbers in parentheses are the indices of singular coordinates that control the open/close of the channel. **(c, d)** Direction of motion for singular coordinates u_0 and u_{41}. The crystal structure is colored **gray**; coordinates with coefficients > 0.1 or < -0.1 are colored **red** (positive) and **blue** (negative). The structure after displacement along u_i is colored **cyan**. **Red** arrows indicate direction of displacement.

Task partitioning among singular coordinates. The clear separation between PEFs of the functional modes and the dissipation modes indicates that they serve distinct functional purposes. As shown in Fig. 6 and supplementary videos, a functional mode involves large-scale rigid body movements of the helices, with the amplitude proportional to its PEF. These movements are driven by small changes (up to 0.45°) in individual backbone dihedrals that have high PEFs (the red loops in Fig. 3a). They appear to be the breathing motion that was a central topic of Mb dynamics [37].

The breathing motion of Mb has two important effects: 1) change R_g of Mb, 2) control ligand migration through the internal cavities (distal pocket and xenon cavities) of Mb [38]. Recent SAXS experiment by Levantino et al found an oscillation of R_g on a picosecond time scale after photolysis [27]. We computed the change in R_g due to change in Mb structure induced by displacements along each functional mode. For each mode, we use the crystal structure S_c as the reference point and obtain new structures $S_i = S_c + \Delta u_i$, with $\Delta u_i = \pm 2.3°$, and then calculate the R_g of S_i. Figure 6a shows that functional modes u_0 and u_3 can change R_g on a similar scale as observed by Levantino et al.

Ligand migration has been the subject of many time-dependent Laue crystallography and computational studies [37, 39-42]. However, it remains an open question what are the protein motions that control ligand migration between these cavities, which are disconnected from each other in the crystal structure. As shown in Fig. 6b and Table S2 [43], a functional mode can merge two or more cavities, enabling the ligand to migrate between these cavities. In this way, a functional mode acts as the gate to the channel connecting two or more cavities. Because motion along a functional mode can open or shut this gate, it regulates ligand migration process. In addition, functional modes can also modulate the volume of individual cavities, adding a further layer of control on ligand migration. Previous computational studies mapped the pathways of ligand migration, but they did not provide information on the specific motions that control these pathways. We showed how these pathways can be easily regulated by a small number of collective modes.

Another long-standing question on Mb dynamics is the coupling between heme vibrations and global motions of the protein matrix. Recent experiment on the ultrafast dynamics of Mb after photolysis using femtosecond crystallography strongly suggested existence of such coupling [26], but does not answer two critical questions: 1) What kind of global motions are coupled to heme vibrations? 2) How are they coupled together? To answer these questions, we identified singular coordinates in which heme dihedrals and bond angles with high PEFs (red curves in Fig. 2b, 2c) have the largest coefficients. These coordinates are dissipation modes with relative high PEFs (>1,000 kJ/mol). Among these four modes, u_{18} and u_{19} couples stretching motions of heme with small-amplitude rigid-body motion of all the long helices (i.e. helices A to H), driven by the loops connecting helices D and E, F and G, G and H. The coupling is mainly between the bond angles of heme and backbone dihedrals and bond angles of these loops, as shown in Fig. S2 and videos S3. On the other hand, u_{39} and u_{41} couple doming and twisting motions of the heme plane with rigid-body motion of helices A, B, E, F, driven by loops connecting C and D, E and F, G and H. The coupling is mainly between heme dihedrals and backbone dihedrals and bond angles of these loops and helix F (Fig. 6d). These modes are the representatives of the coupling between heme vibrations and global protein motions, as there are many other dissipation modes with smaller amplitudes that directly couple heme vibrations with backbone dihedrals and bond angles.

Discussions

Compared to the large-scale conformational dynamics involved in ligand binding and allosteric transitions, the structural dynamics during a protein quake has small amplitude and is subtle. Even though it is fundamentally different from thermal fluctuations, it is a challenge to discern it from the latter. To uncover this intrinsic order buried under thermal noise, a precise method is critical. The energy flow theory allows us to map out the exact energy cost of the motion of every single coordinate during a protein quake, providing a comprehensive way to eliminate noise by proper ensemble average. This enabled us to uncover the detailed molecular mechanism of the energy transduction and structural dynamics of Mb induced by photolysis.

Contrary to the previous conclusion that this process is dominated by diffusion [17, 19, 30], we found that it is carried out by a well-organized hierarchy of energy flow channels synchronized by systematic structural dynamics that are encapsulated by the singular coordinates. The functional modes loosen up the protein to get the dissipation modes going so that energy waves can propagate. The excess energy of the heme is carried by this energy wave across the protein matrix and into the solvents. This mechanism is much more efficient than diffusion because waves are directional and systematic while diffusion is random.

Our results suggest that, in a functional process, the behavior of a protein is more like a fine-tuned mechanical machine designed for a specific purpose, rather than a stochastic system that relies on chance to achieve functional consequences. It is most tempting to speculate that functional processes utilize the same set of modes but excite them to different amplitudes. While energy or signal transduction utilizes synchronized energy waves through many modes, an activated process selectively excites one or a few functional modes to a very high level.

Simulation Method

All simulations were performed using the molecular dynamics software suite GROMACS 2019.2 with CHARMM 36m force field [44-46]. The simulation protocol closely follows that of ref. [19]. The sperm whale Mb molecule used in our simulation is an intermediate state after photodissociation of CO obtained at liquid helium temperature, with the photo-dissociated CO lying on top of the heme pyrrole ring [47]. The Mb molecule was solvated in a water box of $76\ \text{Å} \times 76\ \text{Å} \times 76\text{Å}$. The

simulation system consists of 41,267 atoms in total. The temperature was slowly increased to 300 K, followed by 40 ps of constant temperature molecular dynamics for equilibration [19]. Afterwards, we run a long constant NVE MD simulation for 40 ns. From this trajectory, we select a phase-space condition every 5 ps as the initial condition for the subsequent simulations of the heme cooling process, resulting in 8,000 initial conditions.

To mimic the photolyzed state, an extra 88 kcal/mol = 368 kJ/mol of kinetic energy is then deposited into the heme, following the protocol by Sagnella et al. This excess energy is also similar to the 81 kcal/mol deposited into the heme in the first MD simulation of the cooling of heme after photolysis or photo excitation conducted by Henry, Eaton and Hochstrasser. Addition of this excess kinetic energy amounts to increasing the heme temperature by 400K, which can be achieved by evenly scaling the velocities of heme atoms, for the following reason. Due to fast intramolecular vibrational energy redistribution within the heme, the excess energy is equipartitioned among heme atoms and the initial velocities of heme atoms should follow the Maxwell-Boltzmann distribution at 700 K. This can be achieved by evenly scaling the velocities of heme atoms that follow Maxwell-Boltzmann distribution at 300 K, which is exactly the distribution of the heme velocities in the initial conditions prepared above.

From each initial condition after the rescaling of velocities of heme atoms, a trajectory was launched to simulating the heme cooling after photolysis for 5 ps. The simulations use velocity Verlet algorithm with a 1 fs time step. This process produces 8,000 trajectories for the heme cooling process. The energy flow results reported in the main text were conducted on these trajectories.

Acknowledgements

We thank NSF (award: CHE-1665104) and NIH (R01 GM086536) for financial support and the anonymous reviewers for their excellent suggestions that helped to improve the manuscript.

Supporting Information

Supporting Information contains additional figures, computational details and videos. This information is available free of charge via the Internet at http://pubs.acs.org

References

1. Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E.; Sauke, T. B.; Shyamsunder, E.; Young, R. D., Protein states and proteinquakes. *Proc Natl Acad Sci U S A* **1985,** *82* (15), 5000-4.
2. Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C., Dynamics of Ligand-Binding to Myoglobin. *Biochemistry* **1975,** *14* (24), 5355-5373.
3. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G., The energy landscapes and motions of proteins. *Science* **1991,** *254* (5038), 1598-603.
4. Boehr, D. D.; Dyson, H. J.; Wright, P. E., An NMR perspective on enzyme dynamics. *Chemical Reviews* **2006,** *106* (8), 3055-3079.
5. Boehr, D. D.; Nussinov, R.; Wright, P. E., The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* **2009,** *5* (11), 789-796.
6. Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D., A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007,** *450* (7171), 913-U27.
7. Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D., Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007,** *450* (7171), 838-U13.
8. Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W., Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc Natl Acad Sci U S A* **2003,** *100* (15), 8615-7.
9. Greene, B. I.; Hochstrasser, R. M.; Weisman, R. B.; Eaton, W. A., Spectroscopic studies of oxy- and carbonmonoxyhemoglobin after pulsed optical excitation. *Proc Natl Acad Sci U S A* **1978,** *75* (11), 5255-9.
10. Lim, M.; Jackson, T. A.; Anfinrud, P. A., Nonexponential protein relaxation: dynamics of conformational change in myoglobin. *Proc Natl Acad Sci U S A* **1993,** *90* (12), 5801-4.
11. Mizutani, Y.; Kitagawa, T., Direct observation of cooling of heme upon photodissociation of carbonmonoxy myoglobin. *Science* **1997,** *278* (5337), 443-6.
12. Lim, M.; Jackson, T. A.; Anfinrud, P. A., Binding of CO to myoglobin from a heme pocket docking site to form nearly linear Fe-C-O. *Science* **1995,** *269* (5226), 962-6.
13. Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A., Conformational relaxation and ligand binding in myoglobin. *Biochemistry* **1994,** *33* (17), 5128-45.
14. Sato, A.; Gao, Y.; Kitagawa, T.; Mizutani, Y., Primary protein response after ligand photodissociation in carbonmonoxy myoglobin. *Proc Natl Acad Sci U S A* **2007,** *104* (23), 9627-32.
15. Lim, M. H.; Jackson, T. A.; Anfinrud, P. A., Femtosecond near-IR absorbance study of photoexcited myoglobin: Dynamics of electronic and thermal relaxation. *J Phys Chem-Us* **1996,** *100* (29), 12043-12051.
16. Lian, T. Q.; Locke, B.; Kholodenko, Y.; Hochstrasser, R. M., Energy-Flow from Solute to Solvent Probed by Femtosecond Ir Spectroscopy - Malachite Green and Heme Protein Solutions. *J Phys Chem-Us* **1994,** *98* (45), 11648-11656.
17. Henry, E. R.; Eaton, W. A.; Hochstrasser, R. M., Molecular dynamics simulations of cooling in laser-excited heme proteins. *Proc Natl Acad Sci U S A* **1986,** *83* (23), 8982-6.
18. Kuczera, K.; Lambry, J. C.; Martin, J. L.; Karplus, M., Nonexponential relaxation after ligand dissociation from myoglobin: a molecular dynamics simulation. *Proc Natl Acad Sci U S A* **1993,** *90* (12), 5805-7.

19. Sagnella, D. E.; Straub, J. E., Directed energy "Funneling" mechanism for heme cooling following ligand photolysis or direct excitation in solvated carbonmonoxy myoglobin. *Journal of Physical Chemistry B* **2001,** *105* (29), 7057-7063.
20. Bu, L. T.; Straub, J. E., Simulating vibrational energy flow in proteins: Relaxation rate and mechanism for heme cooling in cytochrome c. *Journal of Physical Chemistry B* **2003,** *107* (44), 12339-12345.
21. Bu, L. T.; Straub, J. E., Vibrational energy relaxation of "tailored" hemes in myoglobin following ligand photolysis supports energy funneling mechanism of heme "cooling". *Journal of Physical Chemistry B* **2003,** *107* (38), 10634-10639.
22. Zhang, Y.; Fujisaki, H.; Straub, J. E., Molecular dynamics study on the solvent dependent heme cooling following ligand photolysis in carbonmonoxy myoglobin. *J Phys Chem B* **2007,** *111* (12), 3243-50.
23. Davarifar, A.; Antoniou, D.; Schwartz, S. D., The Promoting Vibration in Human Heart Lactate Dehydrogenase Is a Preferred Vibrational Channel. *J Phys Chem B* **2011,** *115* (51), 15439-15444.
24. Skopintsev, P.; Ehrenberg, D.; Weinert, T.; James, D.; Kar, R. K.; Johnson, P. J. M.; Ozerov, D.; Furrer, A.; Martiel, I.; Dworkowski, F.; Nass, K.; Knopp, G.; Cirelli, C.; Arrell, C.; Gashi, D.; Mous, S.; Wranik, M.; Gruhl, T.; Kekilli, D.; Brunle, S.; Deupi, X.; Schertler, G. F. X.; Benoit, R. M.; Panneels, V.; Nogly, P.; Schapiro, I.; Milne, C.; Heberle, J.; Standfuss, J., Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature* **2020,** *583* (7815), 314-318.
25. Arnlund, D.; Johansson, L. C.; Wickstrand, C.; Barty, A.; Williams, G. J.; Malmerberg, E.; Davidsson, J.; Milathianaki, D.; DePonte, D. P.; Shoeman, R. L.; Wang, D.; James, D.; Katona, G.; Westenhoff, S.; White, T. A.; Aquila, A.; Bari, S.; Berntsen, P.; Bogan, M.; van Driel, T. B.; Doak, R. B.; Kjaer, K. S.; Frank, M.; Fromme, R.; Grotjohann, I.; Henning, R.; Hunter, M. S.; Kirian, R. A.; Kosheleva, I.; Kupitz, C.; Liang, M.; Martin, A. V.; Nielsen, M. M.; Messerschmidt, M.; Seibert, M. M.; Sjohamn, J.; Stellato, F.; Weierstall, U.; Zatsepin, N. A.; Spence, J. C.; Fromme, P.; Schlichting, I.; Boutet, S.; Groenhof, G.; Chapman, H. N.; Neutze, R., Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat Methods* **2014,** *11* (9), 923-6.
26. Barends, T. R.; Foucar, L.; Ardevol, A.; Nass, K.; Aquila, A.; Botha, S.; Doak, R. B.; Falahati, K.; Hartmann, E.; Hilpert, M.; Heinz, M.; Hoffmann, M. C.; Kofinger, J.; Koglin, J. E.; Kovacsova, G.; Liang, M.; Milathianaki, D.; Lemke, H. T.; Reinstein, J.; Roome, C. M.; Shoeman, R. L.; Williams, G. J.; Burghardt, I.; Hummer, G.; Boutet, S.; Schlichting, I., Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* **2015,** *350* (6259), 445-50.
27. Levantino, M.; Schiro, G.; Lemke, H. T.; Cottone, G.; Glownia, J. M.; Zhu, D.; Chollet, M.; Ihee, H.; Cupane, A.; Cammarata, M., Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat Commun* **2015,** *6*, 6772.
28. Brinkmann, L. U.; Hub, J. S., Ultrafast anisotropic protein quake propagation after CO photodissociation in myoglobin. *Proc Natl Acad Sci U S A* **2016,** *113* (38), 10565-70.
29. Falahati, K.; Tamura, H.; Burghardt, I.; Huix-Rotllant, M., Ultrafast carbon monoxide photolysis and heme spin-crossover in myoglobin via nonadiabatic quantum dynamics. *Nat Commun* **2018,** *9* (1), 4502.
30. Fujisaki, H.; Straub, J. E., Vibrational energy relaxation in proteins. *Proc Natl Acad Sci U S A* **2005,** *102* (19), 6726-31.

31. Leitner, D. M., Energy flow in proteins. *Annu Rev Phys Chem* **2008,** *59*, 233-59.
32. Yu, X.; Leitner, D. M., Anomalous diffusion of vibrational energy in proteins. *Journal of Chemical Physics* **2003,** *119* (23), 12673-12679.
33. Li, W.; Ma, A., Reaction mechanism and reaction coordinates from the viewpoint of energy flow. *J Chem Phys* **2016,** *144* (11), 114103.
34. Li, H.; Ma, A., Kinetic energy flows in activated dynamics of biomolecules. *J. Chem. Phys.* **2020,** *153* (9), 094109.
35. Wu, S.; Li, H.; Ma, A., A Rigorous Method for Identifying One-Dimensional Reaction Coordinate in Complex Molecules. *J. Chem. Theo. Comp.* **2022,** *18* (5), 2836-2844.
36. Wu, S.; Ma, A., Mechanism for the rare fluctuation that powers protein conformational change. *J. Chem. Phys.* **2022,** *156*, 05419.
37. Tomita, A.; Sato, T.; Ichiyanagi, K.; Nozawa, S.; Ichikawa, H.; Chollet, M.; Kawai, F.; Park, S. Y.; Tsuduki, T.; Yamato, T.; Koshihara, S. Y.; Adachi, S., Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. *Proc Natl Acad Sci U S A* **2009,** *106* (8), 2612-6.
38. Tilton, R. F.; Kuntz, I. D.; Petsko, G. A., Cavities in Proteins - Structure of a Metmyoglobin-Xenon Complex Solved to 1.9-A. *Biochemistry* **1984,** *23* (13), 2849-2857.
39. Hummer, G.; Schotte, F.; Anfinrud, P. A., Unveiling functional protein motions with picosecond x-ray crystallography and molecular dynamics simulations. *Proc Natl Acad Sci U S A* **2004,** *101* (43), 15330-4.
40. Case, D. A.; Karplus, M., Dynamics of ligand binding to heme proteins. *J Mol Biol* **1979,** *132* (3), 343-68.
41. Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V., Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *P Natl Acad Sci USA* **2008,** *105* (27), 9204-9209.
42. Maragliano, L.; Cottone, G.; Ciccotti, G.; Vanden-Eijnden, E., Mapping the network of pathways of CO diffusion in myoglobin. *J Am Chem Soc* **2010,** *132* (3), 1010-7.
43. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J., CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **2018,** *46* (W1), W363-W367.
44. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr., 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. *J Chem Theory Comput* **2012,** *8* (9), 3257-3273.
45. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **2008,** *4* (3), 435-447.
46. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017,** *14* (1), 71-73.
47. Schlichting, I.; Berendzen, J.; Phillips, G. N., Jr.; Sweet, R. M., Crystal structure of photolysed carbonmonoxy-myoglobin. *Nature* **1994,** *371* (6500), 808-12.

Graphic TOC Entry