An efficient timer and sizer of biomacromolecular motions

Justin Chan[1,2]¶, Hong-Rui Lin[1,¶], Kazuhiro Takemura[3,¶], Kai-Chun Chang[4], Yuan-Yu Chang[1], Yasumasa Joti[5], Akio Kitao[3] and Lee-Wei Yang [1,6*]*

1. Institute of Bioinformatics and Structural Biology, National Tsing Hua Univ., Taiwan

2. Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Sciences, Academia Sinica, Taiwan

3. School of Life Science and Technology, Tokyo Institute of Technology, Japan

4. Institute of Molecular and Cellular Biology, National Taiwan University, Taiwan

5. XFEL Utilization Division, Japan Synchrotron Research Institute, Japan

6. Physics Division, National Center for Theoretical Sciences, Taiwan

¶Authors contributed equally to this work

*Corresponding authors:

Lee-Wei Yang

Tel: +88635742467;

Fax +88635715934;

Email: lwyang@life.nthu.edu.tw

Institute of Bioinformatics and Structural Biology, National Tsing Hua Univ.,

No. 101, Section 2, Kuang-Fu Road, Hsinchu 30013, Taiwan

Akio Kitao

Tel: +81-(0)3-5734-3373;

Fax: +81-(0)3-5734-3372;

Email: akitao@bio.titech.ac.jp

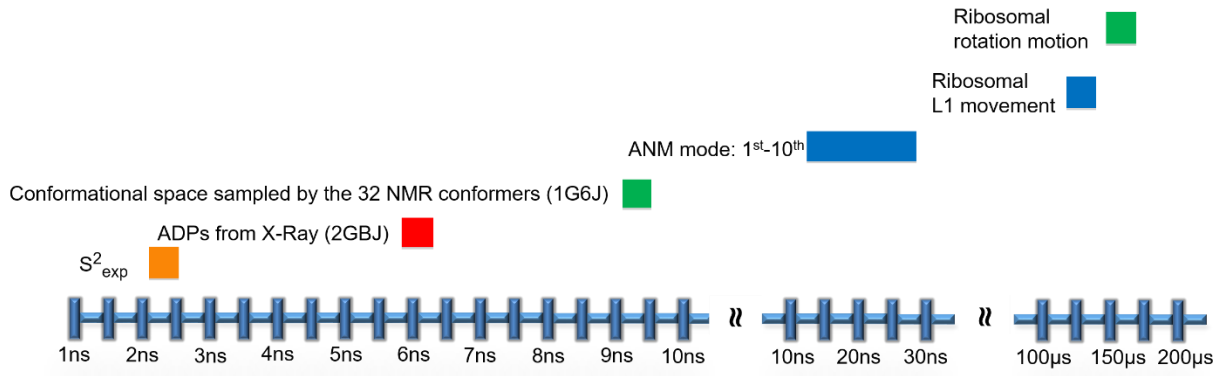
School of Life Science and Technology, Tokyo Institute of Technology,

M6-13, 2-12-1 Ookayama, Meguro, Tokyo 152-8550, JAPAN

Abstract

Life ticks as fast as how efficiently proteins perform their functional dynamics. Well-folded/structured biomacromolecules perform functions via large-scale intrinsic motions across multiple conformational states, which occur at timescales of nano- to milliseconds. Computationally expensive molecular dynamics (MD) simulation has been the only theoretical tool to gauge the time and sizes of these motions, though barely to their slowest ends. Here, we convert a computationally cheap elastic network model (ENM) into a molecular timer and sizer to gauge the slowest functional motions of proteins and ribosome. Quasi-harmonic analysis, fluctuation-profile matching (FPM) and the Wiener–Khintchine theorem (WKT) are used to define the "time-periods", *t*, for anharmonic principal components (PCs) which are validated by NMR order parameters. The PCs with their respective "time-periods" are mapped to the eigenvalues (λ_{ENM}) of the corresponding ENM modes. Thus, the power laws t(ns) = $86.9\lambda_{ENM}^{-1.9}$ and σ^2(Å^2) = $46.1\lambda_{ENM}^{-2.5}$ are established allowing the characterization of the time scales of NMR-resolved conformers, crystallographic anisotropic displacement parameters, and important ribosomal motions, as well as motional sizes of the latter.

TOC Graphics



Keywords

protein dynamics, molecular dynamics simulations, principal component analysis (PCA), elastic network model (ENM), Wiener–Khintchine theorem, Einstein–Smoluchowski relation, time-correlation function, power spectrum, anharmonicity, order parameter, anisotropic displacement parameters, ribosome, ubiquitin

Introduction

The magnitude of protein functional motions for folded proteins (or 'equilibrated' proteins) modulate the rates of their underlying physiological processes[1]. In principle, the wider the conformational spread is, the longer the time it takes. Such a relation is readily realized by analyses of a short molecular dynamics (MD) trajectory. Despite the tremendous software/hardware progress made to accelerate the simulations[2–6], it still requires weeks, if not months, to unravel the conformational spreads and time scales of fully solvated proteins/complexes with the hope that relevant functional motions can be resolved within the length of the simulations.

On the other hand, computationally cheap (taking seconds to minutes) elastic network models (ENMs) have been widely used to study the relative span (variance) of spatial distributions of atoms and predict conformational changes for two decades[7–10]. These physically intuitive models have been used to study vibrational dynamics of all the proteins/supramolecules[11,12] in Protein Data Bank (PDB) where the "directions" of proteins' conformational changes as observed by x-ray crystallography have been satisfactorily reproduced in hundreds of applications and in database-wide studies[13,14] (see the caption in **Figure 1**). However, the absolute time scales and variances of functional modes cannot be directly assessed by ENM due to its harmonic approximation. Functional modes, particularly, involve motions at large scales, wherein proteins traverse across multiple intermediate states (corresponding to local traps in the energy landscape; see the left and middle in **Figure 1**), and therefore are anharmonic in nature[15–17]. In other words, the true challenge lies on how to properly define the "time periods" of these anharmonic modes, which cannot be simply inferred from the length of the simulation (a small protein can travel several times on a

given normal mode in a long MD simulation). Once defined, efficient and accurate methods are needed to predict the "time periods" and absolute variances of these anharmonic modes.

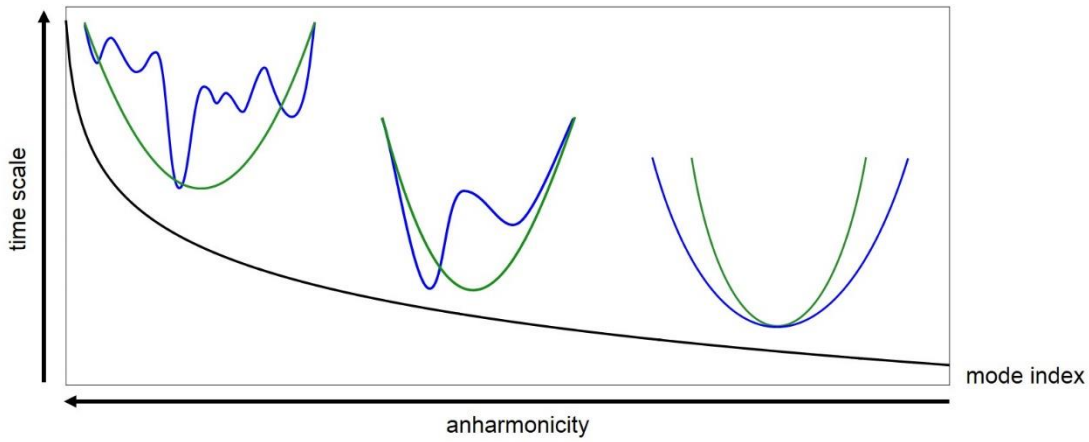


Figure 1. Time scales versus anharmonicity of the modes. The blue and green curves indicate the real and harmonically approximated energy landscapes, respectively. The slower the modes are, the larger the deviations from a simple harmonic approximation of the energy landscape (equivalently stated as having enriched anharmonicity[15–17]). ENM-approximated harmonic energy landscape forms an "envelope" that outlines the real anharmonic energy landscape[14] (to the left in **Figure 1**), which explains the observed correspondence between the theoretically predicted and experimentally characterized "directions" of conformational changes. However, the time scales and absolute spatial span of the modes, among the slowest and often functional, cannot be satisfactorily assessed given the harmonic approximation, which motivated us to design the current method to address the issue. The time scale estimated using the Intensity Weighted Period (IWP) method is introduced in this work (see below).

Below, we give an overview of the new methods for predicting the timescale and size of the dynamics of proteins using MD and ENM. Starting with MD, we perform Principal Component

Analysis (PCA) to decompose the dynamics of the protein into Principal Components (PCs). We then introduce the Intensity Weighted Period (IWP) which we define as the timescale for each PC's motion. Combining the IWP with the Fluctuation Profile Matching (FPM) method, the timescale of the slowest motions captured by experimentally measured dynamical profiles such as NMR root-mean-square-fluctuations (RMSF), anisotropic displacement parameters (ADPs) and NMR order parameters can be predicted, where the last one is used to validate our method. IWP and FPM relies on performing computationally expensive MD simulations. Instead, we use ENM and found two power laws relating each ENM mode's eigenvalue with its timescale and variance, respectively. Finally, we apply the two power laws for analyzing ribosomal motions.

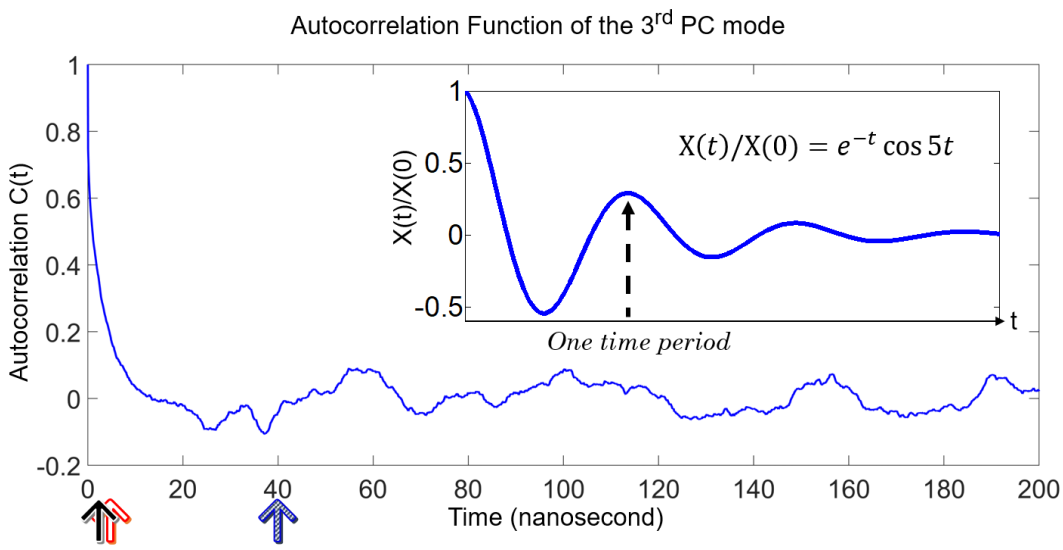


Figure 2. Defining the time scale of the first PC mode using the three methods. The profile of the autocorrelation function for the third PC mode, derived from PCA of a 600-ns simulation of ubiquitin (PDB ID: 1UBQ) and Wiener–Khintchine theorem (**eq 3**), is plotted. Three time constants τ_c = 2.34 ns (indicated by the black solid arrow), τ_r = 2.79 ns (red hollow arrow; obtained using fitting in the time domain) and τ_w = 40.45 ns (blue shaded arrow) that characterize the exponential decay of the autocorrelation function $C(t) = A\exp(-t/\tau_r)$, characteristic time such

that $\tau_c = \sum_{i=1}^{M} C(t_i)\Delta t / C(0)$ and IWP of the power spectrum that is $\tau_w = 2\pi \sum_{i=1}^{M} S(\omega_i)\, \omega_i^{-1} / \sum_{j=1}^{M} S(\omega_j)$, respectively, are used to describe the vibrational relaxation process that mimics a damped harmonic oscillator. In addition, because the Fourier transform of an exponential function is a Lorentzian function[18], we can also obtain τ_r from fitting the entire power spectrum using a Lorentzian function. However, the resulting τ_r is much smaller than τ_c with a fitting correlation close to zero (data not shown). **(Inset)** Theoretical profile of one particle oscillating in an underdamped regime. The equation comparatively shows the equivalent decay and oscillation rate of the autocorrelation function. The vertical arrow indicates the time period, which in this illustrative example is 2π/5.

Results

Overview

Previous analyses of MD trajectories have revealed that the autocorrelation function of the protein's deviation from the mean structure[19] or velocity[15], projected along a certain PC, especially with large and medium variance, appears as a damped harmonic oscillator (**Figure 2**). We tried several methods to properly define the period, or "timescale" in this context, of the damped harmonic oscillator for a PC mode. Among them, the IWP method appears to be the most general and can best reproduce the observed timescale of the MD-sampled oscillatory behavior (**Figure 2**). The IWP is the intensity weighted average of the periods from the power spectrum for a specific PC's autocorrelation function per the Wiener–Khintchine theorem (WKT)[20]. Meanwhile, the FPM method identifies subsets of PCs that can best reproduce (with the highest correlation) the experimentally measured dynamical profile and then assign the IWP of the slowest PC as the timescale of the profile. The subsets of PCs are generated by removing the top-*k* slowest PCs

derived from long MD simulations. This is because we have previously[21] shown that deletion of the slowest ENM modes does not change the correlation between the predicted and experimentally observed B-factors, suggesting that dynamical profiles such as B-factors or anisotropic displacement parameters (ADP) do not include the protein's slowest motions given the confined crystalline environment. In this study, we employ one of the most widely used ENMs, the anisotropic network model (ANM)[10,22,23]. It was known that the mapping of PCA mode and ANM mode is not a one-to-one relationship; a slow PC (ANM) mode is best interpreted by a few slowest ANM (PC) mode[24,25]. FPM provides a robust mapping between a PC mode *k* and an ANM mode *k'* (or an experimental variable) when the variance comprising all the PC modes higher than *k* has the highest correlation with the variance comprising all the ANM modes higher than *k'* (see **Supporting Information** for details). The experimental variables here include NMR order parameters, ADPs from the X-ray structure and RMSF of NMR-solved structures of ubiquitin. To use the ANM as a convenient molecular timer and sizer, we match each ANM mode to a PC mode using FPM. Then, the timescale and variance of an ANM mode are assigned by those of the matched PC mode. We observe that the eigenvalues of ANM modes has a power law relationship with their assigned timescales and variances. Applying the aforementioned procedure to three protein systems, we derive master equations for time and size of protein motions as power laws of ANM eigenvalues. We confirm the size power law by reproducing the observed ribosomal body rotation and estimate the time-scales for ribosomal body rotation and head swiveling motions to a narrower range than what was suggested by MD simulations and experiments.

Determining the time scales of the anharmonic modes

To estimate the true time scale of a PC mode, we first performed a 600-ns MD simulation for the 76 amino acid signaling protein, ubiquitin[26] (PDB ID: 1UBQ). PCA analysis[27] was carried out on the trajectory of the first 72 residues of ubiquitin which are not intrinsically disordered. In short, the covariance matrix, $<\Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}}>$, is constructed using a deviation matrix **Q** and decomposed into its corresponding eigenvalues and eigenvectors such that:

$$< \Delta\mathbf{R}\Delta\mathbf{R}^{\mathrm{T}} > = (M-1)^{-1}\mathbf{Q}\mathbf{Q}^{T} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\boldsymbol{T}} \tag{1}$$

$\Delta\mathbf{R}$ is the $3N$-dimensional deviation vector, and **Q** is the $3N \times M$ matrix, where N and M are, respectively, the number of protein atoms in the analysis and the number of snapshots. Each column in **Q** represents the deviation of a given snapshot from the mean structure, while each element in that column is the deviation of a given atom in the x-, y- or z-dimension. $\mathbf{V}_{3N\times 3N}$ is the eigenvector matrix containing $3N$ eigenvectors (or principal components, PCs), each of which is $3N$-dimensional. $\boldsymbol{\Lambda}$ is the $3N \times 3N$ diagonal matrix of rank-ordered eigenvalues (from large to small).

The snapshots were projected onto all the PCs[15,27] to constitute a projection matrix $\mathbf{U}_{3N\times M}$ as

$$\mathbf{U} = \mathbf{V}^{\boldsymbol{T}}\mathbf{Q} \tag{2}$$

where the row k in **U**, $\mathbf{u}_k = [u_{k,0}, u_{k,1}, \ldots, u_{k,M-1}]$, contains the projections of M snapshots onto a given PC eigenvector $\mathbf{V}_k$ for the k'th PC mode. Each snapshot of the protein structure is a scalar value (PC mode coordinate) on the mode k.

M snapshots projected on mode k together with padded M zeros result in a projection series $u_k(s) = \{u_{k0}, u_{k1}, \dots u_{kM-1}, 0^M, 0^{M+1} \dots 0^{2M-1}\}$, based on which the autocorrelation function for the mode k is calculated as follows. Let $\widetilde{u_k}(\omega) = \{\widetilde{u_{k0}}, \widetilde{u_{k1}}, \dots \widetilde{u_{k2M-1}}\}$ be the Fourier transform of $u_k(s)$, and the spectral density of the process $u_k(s)$ can be defined as $S_k(\omega) = \{|\widetilde{u_{k0}}|^2, |\widetilde{u_{k1}}|^2, \dots |\widetilde{u_{k2M-1}}|^2\}$. According to the Wiener–Khintchine theorem[20], the autocorrelation function of $u_k(s)$, $C_k(t)$, can be calculated as the inverse Fourier transform of the spectral density $S_k(\omega)$ such that

$$C_k(t) = < u_k(s) u_k(t+s) > = \frac{1}{2M} \sum_{f=0}^{2M-1} S_k(f)\, e^{\frac{2\pi i f t}{2M}} \tag{3}$$

where, $f = \omega/2\pi$ is the frequency. The mathematical and computational implementation details are provided in the **Supporting Information**.

Figure 2 shows that the $C_3(t)$ of the third PC mode, where $C_3(0)$ is unity (after normalization) at time zero, drops to where there are negative correlations and rises back above zero before eventually vanishing in an oscillatory manner. Most of the slowest modes demonstrate such a behavior, which can be approximated as a harmonic oscillator in underdamped regimes (illustrated in the inset of **Figure 2**). Our purpose here is to define a reasonable "time period" for such a process, which will be inferred as *the time scale (or "one period") of mode k.*

We initiated three attempts to define the time period of the PC mode. For the first one, we approximated the relaxation of $C(t)$ from unity to zero (the first time it hits a value barely larger than zero) as an exponential decay with a relaxation time constant of τ_r. Because the Fourier transform of an exponential function is a Lorentzian function[18], we can also obtain τ_r from fitting

the entire power spectrum using a Lorentzian function, in contrast with fitting $C(t)$ within a limited time range [when $C(t) > 0$]. For the second method, we define the characteristic time constant τ_c, as the integration of $C(t)$ over the simulation time divided by $C(0)$ which is equal to τ_r if $C(t)$ is indeed an exponentially decaying function. For the third attempt, we simply take the intensity-weighted average of the periods in the power spectrum of $S(\omega)$ such that:

$$\tau_{w,k} = \frac{2\pi \sum_{i=1}^{M} S_k(\omega_i)\omega_i{}^{-1}}{\sum_{i=1}^{M} S_k(\omega_i)} \tag{4}$$

where $S_k(\omega_i)$ is the intensity (or "weight") for the frequency component ω_i (ω_i^{-1} is its corresponding period). As shown in **Figure 2**, τ_w, the intensity-weighted period (IWP), reaches a time scale where the $C(t)$ is close to the peak of the first wave, resembling a time period of a damped oscillator, while neither relaxation time τ_r (obtained from the fitting in both time and frequency domains) nor characteristic time τ_c covers the time span for a period of an approximated damped oscillator. Other slow modes behave similarly to the PC mode 3, while in the fast modes, τ_w, τ_r and τ_c characterize a very fast decay in $C(t)$ and have similar values. The comparison of time scales determined using these three methods for all PC modes are shown in **Figure S1**. Therefore, we choose τ_w to describe the time scale of a PC mode derived from the MD trajectories.

Determining the time scales of experimentally observed dynamical variables

The spatial distribution of every residue (or every atom) in a protein near its folded (equilibrated) state can be observed by Cα RMSF of a NMR-resolved structural ensemble[27], NMR order parameters[28] and x-ray crystallographical anisotropic displacement parameters (ADPs)[29]. Plotting the dynamical value for each residue against its residue index results in so-called "observed fluctuation profiles" (oFPs) (**Figure 3, top,** black circles). Every residue has one magnitude of RMSF, one order parameter and 6 ADP values (including variance xx-, yy-, zz- and covariance

xy-, yz- and xz- components[29]). Concurrently, we derive the theoretical counterparts of these measured variables from a MD snapshot-derived covariance matrix, $<\Delta\mathbf{R}\Delta\mathbf{R}^T>_{MD,k}$, comprising all the PC modes ≥ k^{th} mode (lower PC modes have larger variance)

$$< \Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}} >_{MD,k} = \sum_{a=k}^{3N-6} \lambda_a \mathbf{V}_a \mathbf{V}_a{}^{\boldsymbol{T}} \tag{5}$$

where $\mathbf{V}_a$ is the eigenvector with its corresponding eigenvalue, λ_a, taken from the diagonal of $\mathbf{\Lambda}$ in **eq 1** for the *a*'th PC mode.

Next, we use FPM to find the *k* that can best reproduce the oFPs of interest, we seek the optimal k where the profile derived from $<\Delta\mathbf{R}\Delta\mathbf{R}^T>_{MD,k}$ has the best agreement with the oFPs by having the highest Pearson's correlation coefficient (σ_p)[30]. After finding the best *k* via fluctuation profile matching (FPM), $\tau_{w,k}$ of the PC mode *k* is assigned to the observed dynamical variable, as its estimated time scale.For instance, we can take an ensemble of 32 NMR-determined conformers (PDB ID: 1G6J) and form a Cα-only $3N\times3N$ covariance matrix (see **Supporting Information**). Then, we sum the x-, y- and z-components for each residue from the diagonal in the covariance matrix to obtain an RMSF profile shown in **Figure 3a, top**. There are 210 ($3N - 6 = 210$) $\text{RMSF}_{MD,k}$ profiles, each calculated using all the PC modes ≥ k^{th} mode. Among the 210 RMSF profiles, $\text{RMSF}_{MD,k=25}$ has the highest correlation (0.69) with RMSF_{exp}. Thereby, we claim that the spatial spread of this NMR ensemble takes place within the time scale of about $\tau_{w,k=25}$ = 9.32 ns (**Figure 3a, middle and bottom**).

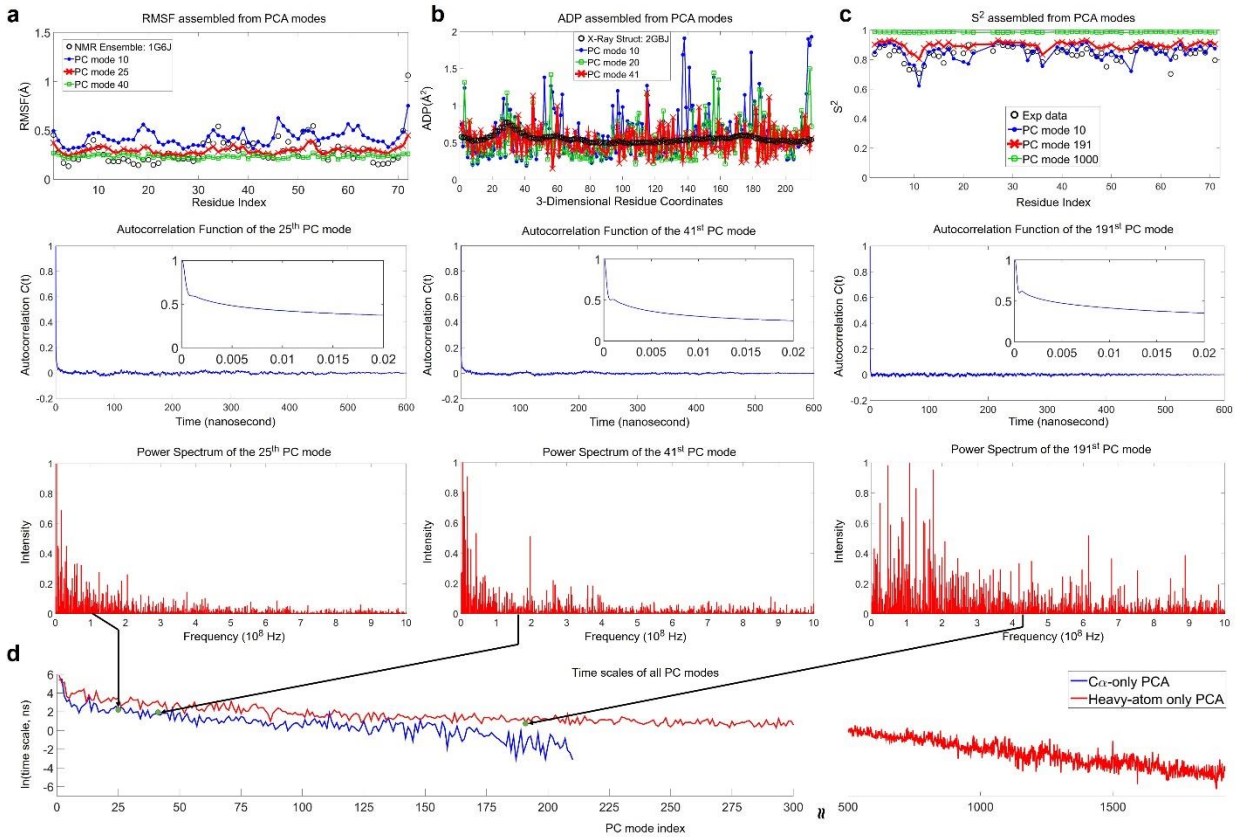


Figure 3. Time scale mapping for different experimental data that characterize protein dynamics. The first rows for **(a-c)** are, respectively, the oFP of RMSF derived from an NMR-determined ensemble (PDB: 1G6J), ADP values (only the diagonal elements are shown) and NMR-measured order parameters against residue index of ubiquitin, where the experimental profiles are plotted in scattered black circles, and the best fit PC mode profile is plotted in thick red lines. **(a, top)** The experimental profile best matches ($\sigma_p = 0.69$) the theoretical RMSF derived from a 600-ns MD simulation, where all PC modes ≥25 are included to contribute to $\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \rangle_{MD,k}$ ($k = 25$); further removing modes higher than mode 24 would impair the correlation. Profiles of different k values are also plotted for comparison. **(a, middle)** The autocorrelation function for the best-matched mode $k = 25$ is computed via WKT following **eq 3**. In addition, **(a, bottom)** its corresponding power spectrum, $S(\omega)$, plotted against its constituent frequencies $f = \omega/2\pi$, can

be obtained before the inverse Fourier transform. Only the contributions of frequencies lower than 10^9 Hz are plotted. A black arrow is drawn to point from the implied IWP's frequency, ($1/\tau_w$), to its corresponding IWP (τ_w) in the ordinate (taken natural logarithm) of the chart in panel **(d)**. **(b and c)** In a similar spirit as **(a)**, panels **(b and c)** are plotted for the experimental ADPs (PDB: 2GBJ) and NMR order parameters[28] of ubiquitin, with the best matched modes determined as k=41 and k=191, respectively. Note that the PCAs performed for **(a and b)** are Cα-based and for **(c)**, the PCA is based on all-heavy-atoms and H atoms bonded to backbone N's. **(d)** IWP (τ_w) is plotted against the PC mode index for both Cα-based ($3N - 6 = 210$ modes; in blue) and all-heavy-atom based ($3N - 6 = 1923$ modes; in red) PCA.

For the case of ADP, ADP_{exp} (taken from PDB ID: 2GBJ) has the highest correlation (0.86) with $ADP_{MD,k=41}$, see **Figure 3a, top**. PC mode 41 has a τ_w value of 6.17 ns, which we assign as the time scale of the ADP distributions for ubiquitin. The correlation coefficients of RMSF and ADP oFPs compared with the theoretical profiles derived using different k's are shown in **Figure S2a**.

The NMR order parameters, S^2, describes the order of the backbone -NH bond vector $r_{ij} = (x_{ij}, y_{ij}, z_{ij})$ pointing from atom i to j (herein, i is N, and j is H) which can be approximated as[27,31,32]

$$S_{ij}^2 = \frac{3}{2}\begin{pmatrix} < x_{ij}^2 >^2 + < y_{ij}^2 >^2 + < z_{ij}^2 >^2 + \\ 2 < x_{ij}y_{ij} >^2 + 2 < x_{ij}z_{ij} >^2 + 2 < y_{ij}z_{ij} >^2 \end{pmatrix} - \frac{1}{2} \quad (6)$$

Here, the length of r_{ij} is normalized to unity. The angular brackets denote averages taken over the M snapshots; herein M = 6,000,000 or 1,200,000 are for the 600 ns and 120 ns trajectories, respectively. S^2 takes on values between 0, for freely rotated bonds (disordered), and unity, for perfectly "ordered" bonds. PCA is carried out for all the heavy atoms plus the hydrogen atoms in

the amino (–NH) groups along the backbone of ubiquitin on the long 600-ns MD simulations and the associated five-constituent 120-ns sub-trajectories.

Through FPM, it was found that $\tau_{w,k=191}$ = 2.34 ns and $\sigma^2_{k=191}$ = 0.44 Å^2 (**Figure 3c** and see correlation as a function of PC mode index, peaked at the mode 191, in **Figure S2b**), characterizes the time scale and size of motion for the observed order parameter profiles[28]. When we applied the current method to shorter trajectories, it was found among five 120-ns simulations that the S^2_{exp} profile consistently maps to the PC modes with τ_w values between 0.7 ns and 1.0 ns (see one 120-ns result in **Figure S3a**).

Hence, despite different simulation lengths being used, our method consistently reports a time scale of 0.7 ns and 2.3 ns for this set of dynamics variables characterized by the NMR relaxation experiments, consistent with the correlation time of methyl symmetry axis motion (300 to 500 ps)[33] using the extended model-free approach, while the "back-calculated" analysis of spectral densities using MD simulations reported time constants of 0.6, 0.9, and 1.5 ns[34].

We observed that $S^2_{MD,k=1}$ shows slightly larger (less ordered) bond fluctuations than S^2_{exp} (**Figure S3b)**, which implies that within the 600 and 120-ns timeframes, the protein samples a wider conformational space than that sampled during the NMR relaxation experiment.

Using the herein proposed method, which integrates PCA, WKT, IWP and FPM, we investigated the time scales of three types of experimentally determined dynamic variables.

Determination of time scales of the ENM modes and inference of time scales from the ENM eigenvalues

It is cumbersome to repetitively perform long MD simulations every time to characterize the time scales of experimentally observed dynamics variables. Here, we devised a computationally light molecular timer that can estimate the time scales of modes from anisotropic network model (ANM)[10,14]. To realize this, FPM of RMSF profiles derived from ANM modes and from MD PC modes was used to map each ANM mode to a PC mode. The ANM derived covariance matrix, $<\Delta\mathbf{R}\Delta\mathbf{R}^T>_{ANM,l}$, for all ANM modes ≥ l (lower ANM modes have lower frequencies or smaller eigenvalues) was calculated as follows

$$< \Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}} >_{ANM,l} = \frac{k_B T}{\gamma} \sum_{a=l+6}^{3N} \lambda_a^{-1} \mathbf{V}_a \mathbf{V}_a^{\boldsymbol{T}} \qquad (7)$$

where $\boldsymbol{V}_a$ and λ_a are the eigenvector and eigenvalue of the a'th ANM mode, respectively. While, γ is the universal spring constant, k_B is the Boltzmann constant and T is the absolute temperature with $\frac{k_B T}{\gamma}$ set to unity. The $\mathrm{RMSF}_{ANM,l}$ and $\mathrm{ADP}_{ANM,l}$ profiles were computed from the $<\Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}}>_{ANM,l}$ matrix with the same methods used for MD-derived profiles of RMSF and ADP.

FPM is applied to each ANM mode l's $\mathrm{RMSF}_{ANM,l}$ profile to identify the kth PC's $\mathrm{RMSF}_{MD,k}$ profile with the highest correlation. By doing this, every ANM mode l can be assigned the time scale (τ_w) of its best-matched k^{th} PC mode. If there are more than one ANM mode mapped to a PC mode, the ANM mode having the highest correlation with that PC mode would be kept. We then fit the remaining ANM eigenvalues (λ_{ANM}) with their mapped time scales (t_{ANM}) in the form of a power law $t_{ANM} = c \times \lambda_{ANM}^{d}$ to obtain the constants c and d, as described in the **Supporting Information** and **Figure 4a**. The time scales estimated from the power law can then be used to

study the slowest/functional modes (low ANM modes) of proteins and FPM can be used to estimate the time scales of experimentally observed dynamics variables (see below).

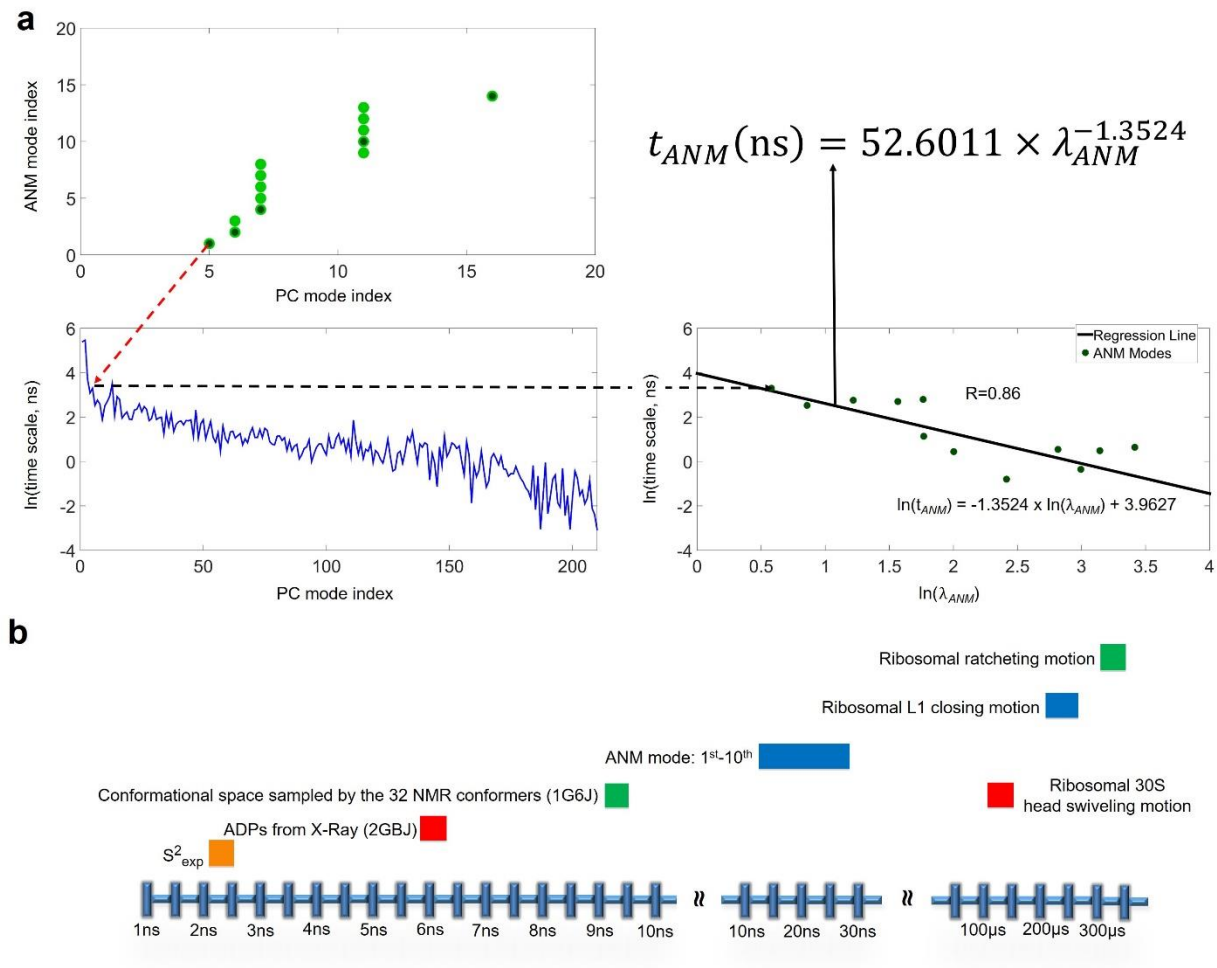


Figure 4. Frequency mapping for ANM to obtain the power law of τ_w as a function of ANM eigenvalues. (a) The ANM analysis and PCA of long MD trajectories are obtained for ubiquitin (residues 1-72, Cα-only). The τ_w of every PC mode obtained using IWP was mapped to the most correlated ANM mode using FPM. **(a, top left),** First *l*-1 ANM modes are removed to obtain a covariance matrix comprising modes ≥*l*. The RMSF of each residue, $\mathrm{RMSF}_{ANM,l}$, is the square root of the sum of x-, y- and z-components of its variance. This $\mathrm{RMSF}_{ANM,l}$ is compared to the

$RMSF_{MD,k}$ comprising PC modes $\geq k$. Each ANM mode l (light-green circles) is mapped to the highest correlated PC mode k by comparing RMSF profiles. When multiple ANM modes are mapped to the same PC mode k, only the pair with the highest correlation is kept (dark-green circles). The τ_w of the PC mode k is then assigned to the ANM mode l. For example, the first (l = 1) ANM mode maps to the 5th PC mode (red arrow). **(a, bottom left),** The τ_w's are plotted against the PC mode index (identical to the blue curve in **Figure 3d**). **(a, bottom right),** The τ_w's are plotted against λ_{ANM} (the eigenvalues of ANM modes). Linear regression gives the power law $t_{ANM}(\text{ns}) = 52.6011 \times \lambda_{ANM}^{-1.3524}$ for ubiquitin. **(b)** Time scales of the examined dynamics variables, including the ANM modes, are assigned by the PCA+WKT+FPM approach for ubiquitin, except for the time scales of ribosomal motions that are estimated by the power law $t_{ANM}(\text{ns}) = 65.4392 \times \lambda_{ANM}^{-2.1386}$ derived from the combined τ_w and λ_{ANM} of ubiquitin and FGF2.

For the case of ubiquitin, ANM mode 1 is mapped to PC mode 5 therefore it is assigned the time scale, $\tau_{w,k=5}$ = 27.03 ns (**Figure 4a**). Similarly, the $RMSF_{ANM}$ for modes 4 and 10 are mapped to PC modes 7 and 11 with the corresponding time periods τ_w of 21.95 and 14.87 ns, respectively. This suggests that the motions described by the first 10 ANM modes for ubiquitin, or the slowest end of ANM, should occur within the timeframe of 13 to 28 ns (see **Figure 4b**). Fitting the λ_{ANM}'s with the t_{ANM}'s we obtain a power law for ubiquitin, $t_{ANM}(\text{ns}) = 52.6011 \times \lambda_{ANM}^{-1.3524}$.

To extend the applicability of this method, we apply the aforementioned method to two other proteins. We performed PCA+IWP analysis on the 200-ns MD trajectories of an FGF2 monomer (126 residues; PDB ID: 1BFG) and HPNAP (144 residues; PDB ID: 1JI4) to derive time scales (τ_w) of each protein's PC modes. Then, we combined the mapped time scales of the ANM modes

and the corresponding eigenvalues for these three proteins (ubiquitin, FGF2 and HPNAP) to derive the general power law $t_{ANM}(\mathrm{ns}) = 86.9387 \times \lambda_{ANM}^{-1.8886}$ (**Figure S4a**). We would like to acknowledge that the general power law derived here from the three proteins of different sizes (length) is nowhere final and presented as a seminal effort for further studies and improvements.

Although the dynamics is apparently a function of protein sizes as well as internal topology, we noted that a correlation can be found between the eigenvalues of the slowest modes and the corresponding protein sizes (**Figure S4b**), which implies that considering eigenvalue alone could have taken into account the protein size effect in our power laws.

In the next section, we will show that good predictions of time scales for many dynamics variables can be obtained using this general power law. Thus, a simple molecular timer that characterizes functional motions of biomolecules is obtained. **Figure 4b** summarizes the time scales of all the discussed experimental observables and ANM modes of interest.

Verification of the general ANM power law via the dynamics analysis of ubiquitin, FGF2 and HPNAP

To validate the applicability of this general power law, we now examine the estimated time scales of the experimental RMSF and ADP profiles (**Figures 3a and 3b**) using ANM instead of PCA of a long MD trajectory. The results showed that theoretical ADPs predicted from removing the first 7 ANM modes agree the best with the experimental ADPs. By substituting the eigenvalue of mode 8 in the general power law, we obtain a time scale of 5.38 ns, which is close to the earlier PCA result of 6.17 ns. In a similar vein, the RMSF profile derived from the 32 NMR conformers

(PDB ID: 1G6J) matches the best with the theoretical RMSF comprising the ANM mode 4 and above, which maps to a time scale of 8.70 ns according to the general power law. This is close to the earlier PCA mapping result of 9.32 ns (mapped to the 25th PC mode). Furthermore, for FGF2 and HPNAP, we found the time scale of each ANM mode predicted by the general power law closely agrees with that estimated from FPM+WKT, with correlations of 0.95 for FGF2 and 0.93 for HPNAP, respectively.

Derive ANM power law for the sizes of functional motions

Since each ANM mode has been mapped to a PC mode, one can describe the PC mode variance (the eigenvalue of PC mode) as a function of the corresponding ANM eigenvalue. Combining the aforementioned three protein cases, the linear regression of a log-log plot provides a new power law such that $\sigma^2(\text{Å}^2) = 46.0538 \times \lambda_{ANM}^{-2.5085}$ (see the **Supporting Information**). This power law, governing the sizes (variance) of motion as a function of ANM eigenvalues, together with the time power law jointly suggest that the protein's functional motions follow a $t^{1.33} \sim \langle\sigma^2\rangle$ relationship. In the meanwhile, the normal mode theory provides a relation $\langle\sigma^2\rangle = k_B T/m\omega^2$ for the harmonic oscillator (where ω is the frequency for the normal mode of interest), suggesting that $t^2 \sim \langle\sigma^2\rangle$, while the Einstein–Smoluchowski relation (or fluctuation-dissipation theorem) for freely diffused particle in 1D describes a $\langle\sigma^2\rangle = 2Dt$ relation, or equivalently $t \sim \langle\sigma^2\rangle$.

Analysis of ribosomal motions

The power laws are applied to analyzing the ribosomal body rotation motion (ratcheting)[11,35] between the 30S and 50S subunits of the *Thermus thermophilus* ribosome[35], which is essential for ribosomal translocation during protein translation[36,37]. We perform ANM analysis on the non-

rotated ribosome conformation and the 25th slowest ANM mode is identified as the ratcheting motion, see details in **Supporting Information**. The deformation vector (eigenvector) for mode 25 is scaled by taking the square root of the variance obtained from the variance power law. The predicted rotated ribosome conformation (**Supporting Information**) generated with the deformation vector for mode 25 shows a 9° rotation which is close to the experimentally measured rotation of 7° (according to the non-rotated state [PDB ID: 4V6F] and rotated state [PDB ID: 4V9H])[38,39].

The estimated timescale of the ribosomal body-rotation (ratcheting) according to simulation[40] and stopped-flow[37,41] is between >1.3 μs and <5.0 ms, respectively. Individual power laws estimated from ubiquitin and FGF2 predicts the timescale to be between 11.5 μs and 651.9 μs, respectively. However, the HPNAP-derived power law estimates a timescale of less than 1.3 μs, inconsistent with the simulation results[40]. Therefore, for predicted time scales longer than 600 ns, which is the longest simulation we conducted to derive the previous power laws, we adopt the power law derived from only the two proteins ubiquitin and FGF2, which gives $t_{ANM}(\mathrm{ns}) = 65.4392 \times \lambda_{ANM}^{-2.1386}$ (see **Figure S5**). With this power law, we obtain a timescale of 327.5 μs for the ribosomal ratcheting motion (see timescale estimates for all the slowest ribosomal normal modes in **Table S1**), which is less than 2-fold longer than 180 μs, the time scale estimated by the power law derived from all the 3 proteins ($t_{ANM}(\mathrm{ns}) = 86.9387 \times \lambda_{ANM}^{-1.8886}$). Using the former power law (for estimated time scale > 600 ns), the L1 stalk closing and 30S head swiveling motion are estimated to have a time scale of 202-253 and 128 μs, respectively (see **Supporting Information)**.

Discussion

In this study, we show that the distributions (or profiles) of dynamical variables can serve as a hallmark to suggest the underlying time scale and variance needed to produce such distributions. Using FPM, the dynamical profile can be mapped to a specific ANM mode whose timescale and variance can be estimated by two tentative power laws established in this study from running long MD simulations, PCA, IWP and FPM.

The two power laws we derived are combined to give $\langle\sigma^2\rangle = 0.1223t^{1.3282}$. Therefore, for the motions (variance) of the same size ($\langle\sigma^2\rangle$), a harmonic oscillator travels the fastest ($t^2 \sim \langle\sigma^2\rangle$)[10], followed by a "guided" (structure-encoded) diffusion motion ($t^{1.33} \sim \langle\sigma^2\rangle$) along functional modes for a solvated protein, and the slowest (the least efficient) goes to the free diffusion ($t \sim \langle\sigma^2\rangle$)[20]. The guided motions take much longer time than what harmonic oscillators would anticipate for particularly large amplitude motions, while the difference between the two is negligible for small conformational changes (fast motions). The result can imply delicate time controls in biological functions at the molecular level. In that, higher-order structures (e.g., translating ribosome) or enzymes central to the signal cascade (e.g., phosphorylation-executing adenylate kinases) would take a much longer time to involve large conformational rearrangements in executing critical functions to ensure that the event is rare and manageable, compared to what is needed for housekeeping proteins, such as lysozyme or ubiquitin, to maintain homeostasis. We expect that the introduced molecular timer and sizer with its accompanying theories can aid in our quantitative understanding of functional and anharmonic motions of biomolecules. The codes for performing the analysis are available on https://github.com/Hong-Rui/bioStructureM.

Methods

The details of the methods of this study can be found in the **Supporting Information**.

MD simulations

Detailed simulation protocols have been previously reported[42]. Briefly, a 600ns MD simulation for Ubiquitin was performed using the PMEMD module of AMBER 10 with the ff99SB force field[43], while CHARMM36[44] force field was used to perform 200ns MD simulations for FGF2 and HPNAP with NAMD 2.10 [45]. The TIP3P model was used, and the distance between the outer most protein atom and the closest simulation box face in the initial setup was 20 Å. All systems were brought to thermodynamic equilibrium at 300 K and 1 atm using a weak coupling thermostat and a barostat. The equations of motion were integrated with a time step of 2 fs/step. The long-range Coulomb energy was evaluated using the particle mesh Ewald (PME) method. MD snapshots were stored at a rate no less than one snapshot per 1 picosecond.

Associated Content

Supporting Information. PDF file containing details of the principal component analysis (PCA), Wiener–Khintchine theorem (WKT), Intensity Weighted Period (IWP), Fluctuation Profile Mapping (FPM), theoretical profiles of RMSF, ADP and order parameters, Anisotropic Network Model (ANM), fitting the time and variance power laws and identification and prediction of the size and time scale of ANM modes corresponding to important ribosomal motions. Figures showing time scales computed using IWP, characteristic time and relaxation time, correlations between the MD derived RMSF, ADP and order parameter profiles using all PC modes $\geq k$ with their respective oFPs of Ubiquitin, computed order parameter from the 120 ns and 600 ns trajectories of Ubiquitin compared to the experimental order parameter, comparison of IWP with the power law estimated time scale, relationship between protein sizes and ANM's slowest mode and collectivity of the slowest 50 ANM modes of the non-rotated ribosome. Table containing the eigenvalues and time power law estimated time scales for the slowest 50 ANM modes of the non-rotated ribosome. 1UBQ.zip, FGF2.zip and HPNAP.zip: input files for performing MD simulations for Ubiquitin, FGF2 and HPNAP respectively.

Author Information

Notes

The authors declare no competing financial interests.

Acknowledgements

We acknowledge the computational resources supported by High Performance Computing Infrastructure (HPCI), Japan, by National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) of Taiwan, and by supercomputers at the RCCS, The National Institute of Natural Science, and ISSP, The University of Tokyo. J.C.

acknowledges financial supports from MOST and Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan. This work is funded by the Ministry of Science and Technology (104-2113-M-007-019 and 106-2313-B-007-001-) and National Center for Theoretical Sciences, Taiwan to L.W.Y and by MEXT/JSPS KAKENHI (No. 25104002 and 15H04357) to A.K., as the "Priority Issue on Post-K Computer" (Building Innovative Drug Discovery Infrastructure Through Functional Control of Biomolecular Systems).

References

(1) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. Dynamically Driven Protein Allostery. *Nat. Struct. Mol. Biol.* **2006**, *13* (9), 831–838.

(2) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542–1555.

(3) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (80-.).* **2010**, *330* (6002), 341–346.

(4) Sanbonmatsu, K. Y.; Joseph, S.; Tung, C.-S. Simulating Movement of TRNA Into the Ribosome During Decoding. *Proc. Natl. Acad. Sci.* **2005**, *102* (44), 15854–15859.

(5) Ratje, A. H.; Loerke, J.; Mikolajka, A.; Brünner, M.; Hildebrand, P. W.; Starosta, A. L.; Dönhöfer, A.; Connell, S. R.; Fucini, P.; Mielke, T.; et al. Head Swivel on the Ribosome Facilitates Translocation by Means of Intra-Subunit TRNA Hybrid Sites. *Nature* **2010**, *468* (7324), 713–716.

(6) Whitford, P. C.; Onuchic, J. N.; Sanbonmatsu, K. Y. Connecting Energy Landscapes with Experimental Rates for Aminoacyl-TRNA Accommodation in the Ribosome. *J. Am. Chem. Soc.* **2010**, *132* (38), 13170–13171.

(7) Flory, P. J.; Gordon, M.; McCrum, N. G. Statistical Thermodynamics of Random Networks [and Discussion]. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1976**, *351* (1666), 351–380.

(8) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77* (9), 1905–1908.

(9) Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys. Rev. Lett.* **1998**, *80* (12), 2733–2736.

(10) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001**, *80* (1), 505–515.

(11) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global Ribosome Motions Revealed with Elastic Network Model. *J. Struct. Biol.* **2004**, *147* (3), 302–314.

(12) Li, H.; Chang, Y.-Y. Y.; Yang, L.-W. W.; Bahar, I. IGNM 2.0: The Gaussian Network Model Database for Biomolecular Structural Dynamics. *Nucleic Acids Res.* **2016**, *44* (D1), D415-22.

(13) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. Global Dynamics of Proteins: Bridging between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39* (39), 23–42.

(14) Li, H.; Chang, Y.-Y.; Lee, J. Y.; Bahar, I.; Yang, L.-W. DynOmics: Dynamics of Structural

Proteome and Beyond. *Nucleic Acids Res.* **2017**, *45* (W1), W374–W380.

(15) Kitao, A.; Hirata, F.; Gō, N. 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.* **1991**, *158* (2–3), 447–472.

(16) Kitao, A.; Go, N. Investigating Protein Dynamics in Collective Coordinate Space. *Curr. Opin. Struct. Biol.* **1999**, *9* (2), 164–169.

(17) Yang, L.-W.; Chng, C.-P. Coarse-Grained Models Reveal Functional Dynamics--I. Elastic Network Models--Theories, Comparisons and Perspectives. *Bioinform. Biol. Insights* **2008**, *2*, 25–45.

(18) Boas, M. L. Mathematical Methods in the Physical Sciences. In *The Journal of Symbolic Logic*; 2006; Vol. 57, pp 340–389.

(19) Okan, O. B.; Atilgan, A. R.; Atilgan, C. Nanosecond Motions in Proteins Impose Bounds on the Timescale Distributions of Local Dynamics. *Biophys. J.* **2009**, *97* (7), 2080–2088.

(20) McQuarrie, D. A. The Time-Correlation Function Formalism II. In *Statistical Mechanics*; University Science Books, 2000; pp 543–592.

(21) Yang, L.-W.; Eyal, E.; Chennubhotla, C.; Jee, J.; Gronenborn, A. M.; Bahar, I. Insights into Equilibrium Dynamics of Proteins from Comparison of NMR and X-Ray Data with Computational Predictions. *Structure* **2007**, *15* (6), 741–749.

(22) Chandrasekaran, A.; Chan, J.; Lim, C.; Yang, L.-W. Protein Dynamics and Contact Topology Reveal Protein-DNA Binding Orientation. *J. Chem. Theory Comput.* **2016**, *12* (11), acs.jctc.6b00688.

(23) Li, H.; Sakuraba, S.; Chandrasekaran, A.; Yang, L.-W. Molecular Binding Sites Are Located near the Interface of Intrinsic Dynamics Domains (IDDs). *J. Chem. Inf. Model.* **2014**, *54* (8), 2275–2285.

(24) Bakan, A.; Bahar, I. The Intrinsic Dynamics of Enzymes Plays a Dominant Role in Determining the Structural Changes Induced upon Inhibitor Binding. *Proc. Natl. Acad. Sci.* **2009**, *106* (34), 14349–14354.

(25) Nicolay, S.; Sanejouand, Y.-H. Functional Modes of Proteins Are among the Most Robust. *Phys. Rev. Lett.* **2006**, *96* (7), 078104.

(26) Vijay-kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194* (3), 531–544.

(27) Yang, L.-W.; Eyal, E.; Bahar, I.; Kitao, A. Principal Component Analysis of Native Ensembles of Biomolecular Structures (PCA_NEST): Insights into Functional Dynamics. *Bioinformatics* **2009**, *25* (5), 606–614.

(28) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin from 15N NMR Relaxation. *J. Am. Chem. Soc.* **1995**, *117* (50), 12562–12566.

(29) Eyal, E.; Chennubhotla, C.; Yang, L.-W.; Bahar, I. Anisotropic Fluctuations of Amino Acids in Protein Structures: Insights from X-Ray Crystallography and Elastic Network Models. *Bioinformatics* **2007**, *23* (13), i175–i184.

(30) Pearson, K. Note on Regression and Inheritance in the Case of Two Parents. *Proc. R. Soc. London* **1895**, *58* (1), 240–242.

(31) Henry, E. R.; Szabo, A. Influence of Vibrational Motion on Solid State Line Shapes and NMR Relaxation. *J. Chem. Phys.* **1985**, *82* (11), 4753.

(32) Best, R. B.; Vendruscolo, M. Determination of Protein Structures Consistent with NMR Order Parameters. *J. Am. Chem. Soc.* **2004**, *126* (26), 8090–8091.

(33) Lee, A. L.; Flynn, P. F.; Wand, A. J. Comparison of 2 H and 13 C NMR Relaxation Techniques for the Study of Protein Methyl Group Dynamics in Solution. *J. Am. Chem. Soc.* **1999**, *121* (12), 2891–2902.

(34) Nederveen, A. J.; Bonvin, A. M. J. J. NMR Relaxation and Internal Dynamics of Ubiquitin from a 0.2 Ms MD Simulation. *J. Chem. Theory Comput.* **2005**, *1* (3), 363–374.

(35) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy. *Proc. Natl. Acad. Sci.* **2003**, *100* (16), 9319–9323.

(36) Frank, J.; Agrawal, R. K. A Ratchet-like Inter-Subunit Reorganization of the Ribosome during Translocation. *Nature* **2000**, *406* (6793), 318–322.

(37) Cornish, P. V.; Ermolenko, D. N.; Noller, H. F.; Ha, T. Spontaneous Intersubunit Rotation in Single Ribosomes. *Mol. Cell* **2008**, *30* (5), 578–588.

(38) Jenner, L. B.; Demeshkina, N.; Yusupova, G.; Yusupov, M. Structural Aspects of Messenger RNA Reading Frame Maintenance by the Ribosome. *Nat. Struct. Mol. Biol.* **2010**, *17* (5), 555–560.

(39) Tourigny, D. S.; Fernandez, I. S.; Kelley, A. C.; Ramakrishnan, V. Elongation Factor G Bound to the Ribosome in an Intermediate State of Translocation. *Science (80-.).* **2013**, *340* (6140), 1235490–1235490.

(40) Whitford, P. C.; Blanchard, S. C.; Cate, J. H. D.; Sanbonmatsu, K. Y. Connecting the Kinetics and Energy Landscape of TRNA Translocation on the Ribosome. *PLoS Comput. Biol.* **2013**, *9* (3), e1003003.

(41) Guo, Z.; Noller, H. F. Rotation of the Head of the 30S Ribosomal Subunit During MRNA Translocation. *Proc. Natl. Acad. Sci.* **2012**, *109* (50), 20391–20394.

(42) Takemura, K.; Kitao, A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. *J. Phys. Chem. B* **2012**, *116* (22), 6279–6287.

(43) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone

Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65* (3), 712–725.

(44) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.

(45) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.