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[3,7] and Lee-Wei Yang [1,2,6,7*]*

[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
[7]Lead Contacts
[¶]Authors contributed equally to this work

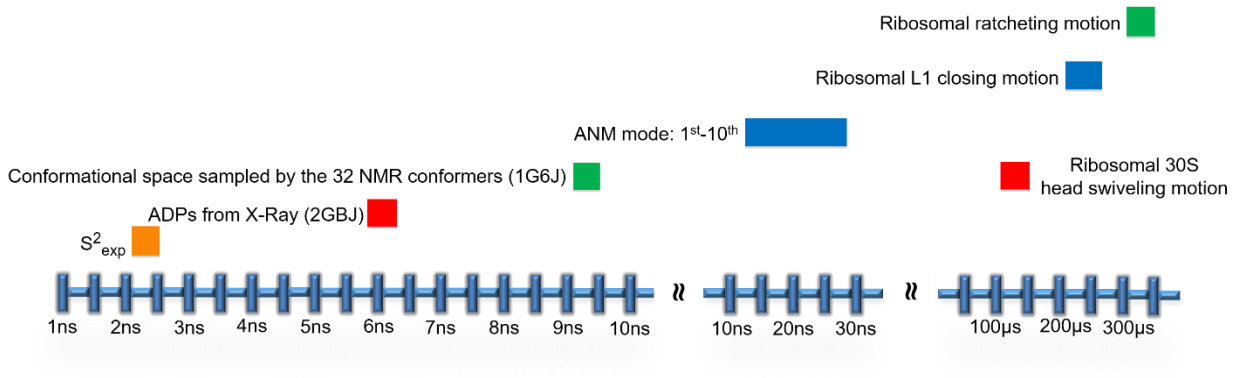
*Correspondence:
Lee-Wei Yang
Tel: +88635742467;
Fax +88635715934;
Email: lwyang@life.nthu.edu.tw
Institute of Bioinformatics and Structural Biology, National Tsing Hua Univ.,
No. 101, Section 2, Kuang-Fu Road, Hsinchu 30013, Taiwan

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

SUMMARY

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.

GRAPHICAL ABSTRACT



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 (Popovych et al., 2006). In principle, the wider the conformational spread is, the longer the time it takes, although the time and space relationship for protein motions damped in water may not be readily described by free diffusion of harmonic oscillators (Atilgan et al., 2001; Bahar et al., 1998; Flory et al., 1976; Li et al., 2016; Tirion, 1996; Wang et al., 2004). Understanding such a relation can bring important insights in protein dynamics including but not limited to the time scales of molecular controls as well as sizes of conformational changes that support the molecular function or facilitate two-body interactions. Molecular dynamics (MD) simulations, though computational costly, has been the only option to explore such a relationship (Götz et al., 2012; Ratje et al., 2010; Sanbonmatsu et al., 2005; Shaw et al., 2010; Whitford et al., 2010), while larger and time-wise lengthier motions (>10s of microseconds for large macromolecules) are still not within the reach of modern MD.

On the other hand, computationally cheap (>5 order cheaper than MD) elastic network models (ENMs) have been widely used for more than two decades (Atilgan et al., 2001; Bahar et al., 1998; Flory et al., 1976; Tirion, 1996) to study and predict the span (variance) of spatial distributions of protein conformations accessible to a molecule. These physically intuitive models have been used to study vibrational dynamics of all the proteins/supramolecules (Li et al., 2016; Wang et al., 2004) 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 (Bahar et al., 2010; Li et al., 2017) (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 (Kitao and Go, 1999; Kitao et al., 1991; Yang and Chng, 2008). 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.

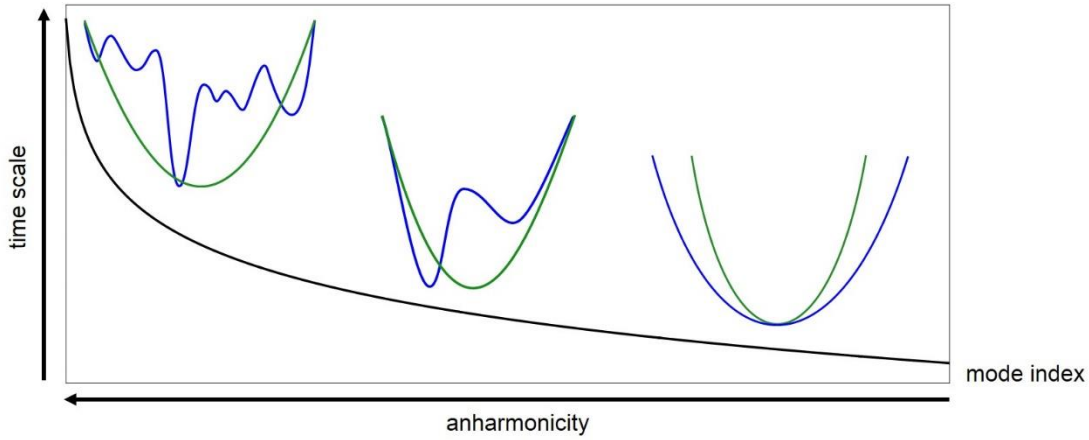


Figure 1. Time scales versus anharmonicity of the modes. The blue and green curves indicate the real and harmonically approximated energy landscapes, respectively. The slower the modes are, the larger the deviations from a simple harmonic approximation of the energy landscape (equivalently stated as having enriched anharmonicity (Kitao and Go, 1999; Kitao et al., 1991; Yang and Chng, 2008)). ENM-approximated harmonic energy landscape forms an "envelope" that outlines the real anharmonic energy landscape (Li et al., 2017) (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).

In this study, we first use Principal Component Analysis (PCA) of MD trajectories, time-correlation functions, Wiener–Khintchine theorem (WKT) (McQuarrie, 2000) to define "time periods" for all the anharmonic and harmonic modes. Then, we use a mapping technique, Fluctuation Profile Matching (FPM), to map a dynamic variable (containing information of spatial spread of protein atoms; see below) or an ENM mode to a Principal Component (PC); hence the time scale of that PC can be assigned to the dynamic variable or ENM mode of interest. When every ENM mode is assigned a time scale or MD-sampled conformational spread (variance), we can then derive corresponding power laws to describe how the time scale and conformational spread of a normal mode can be functions of its eigenvalue.

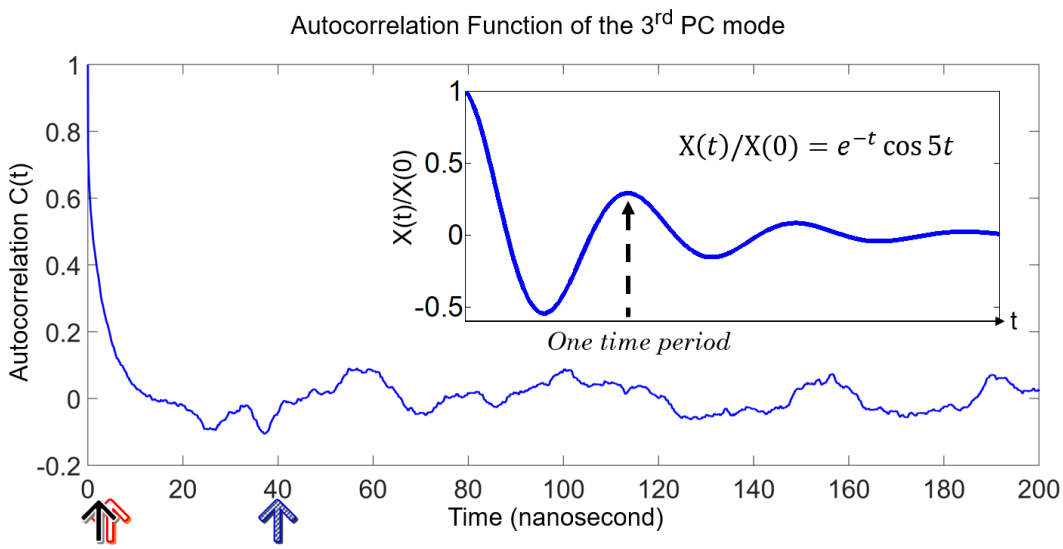


Figure 2. Defining the time scale of a PC mode using three methods. The profile of the autocorrelation function for the third PC mode, derived from PCA of a 600-ns simulation of ubiquitin (PDB ID: 1UBQ) and Wiener–Khintchine theorem (**Equation 1**), is plotted. Three time constants τ_c = 2.34 ns (indicated 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 (Boas, 2006), 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 $2\pi/5$.

RESULTS

Overview

As a methodological overview in finer detail, **Figure 2** describes the autocorrelation function of a PC mode of ubiquitin. As previously described in either position (Okan et al., 2009) or velocity (Kitao et al., 1991) autocorrelation functions, the function decays exponentially as a damped harmonic oscillator (**Figure 2**). We tried three methods to define the period, or "timescale" in this context, of this damped harmonic oscillator for a PC mode. Among them, the Intensity Weighted Period (IWP; see definition below) requires the least assumption and appears to best describe the observed oscillatory behavior (**Figure 2**). Next, FPM identifies subsets of PCs that best reproduce (per the highest correlation) the experimentally observed profile of a given dynamical variable (including root-mean-square-fluctuations (RMSF) of a structural ensemble (Yang et al., 2009), anisotropic displacement parameters (ADPs) (Eyal et al., 2007) and NMR order parameters (Tjandra et al., 1995)) and assign the IWP (or "time scale") of the slowest PC as the timescale of that 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 (Yang et al., 2007) 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 (Yang et al., 2007). In this study, we employ one of the most widely used ENMs, the anisotropic network model (ANM) (Atilgan et al., 2001; Chandrasekaran et al., 2016; Li et al., 2014). Note that it was known that the mapping of PCA mode and ANM

mode (or other ENM mode) is not a one-to-one relationship; a slow PC (ANM) mode is best interpreted by a few slowest ANM (PC) mode (Bakan and Bahar, 2009; Nicolay and Sanejouand, 2006). 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 **STAR Methods** for FPM's mathematical definition). The experimental variables here include NMR order parameters, ADPs from the X-ray structure and RMSF of NMR-solved structural ensemble 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 the timescale and variance 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, ubiquitin, FGF2 and HP-NAP, 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 time range than those 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 (Vijay-kumar et al., 1987) (PDB ID: 1UBQ). PCA analysis (Yang et al., 2009) was carried out on the trajectory of the first 72 residues of ubiquitin, which are structurally ordered to derive the covariance matrix $<\Delta\mathbf{R}\Delta\mathbf{R}^{T}>$ and the projection of each snapshot to each of the eigenvectors (see details in the **STAR Methods**). In other words, a

snapshot is a scalar value (PC mode coordinate) on the mode k., $\mathbf{V}_k$. The total M snapshots projected on mode k together with padded M zeros result in a projection series $u_k(s) = \{u_{k0}, u_{k1}, \dots u_{kM-1}, 0^M, 0^{M+1} \dots 0^{2M-1}\}$, where u_{kj} means the j-th snapshot projected on the mode k, 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 (McQuarrie, 2000), 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

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

(Equation 1)

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

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 (Boas, 2006), 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:

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

(Equation 2)

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.

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 (Yang et al., 2009), NMR order parameters (Tjandra et al., 1995) and x-ray crystallographical anisotropic displacement

parameters (ADPs) (Eyal et al., 2007). 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 (Eyal et al., 2007)). 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 $\geq k^{th}$ mode (lower PC modes have larger variance)

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

(Equation 3)

where $\mathbf{V}_a$ is the eigenvector with its corresponding eigenvalue, λ_a, taken from the diagonal of $\boldsymbol{\Lambda}$ in **Equation S1** in the **STAR Methods** for the *a*'th PC mode; *N* is the total number of heavy atoms in the protein.

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) (Pearson, 1895). 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 (see **Figure 3** and **Supplemental Information** for examples).

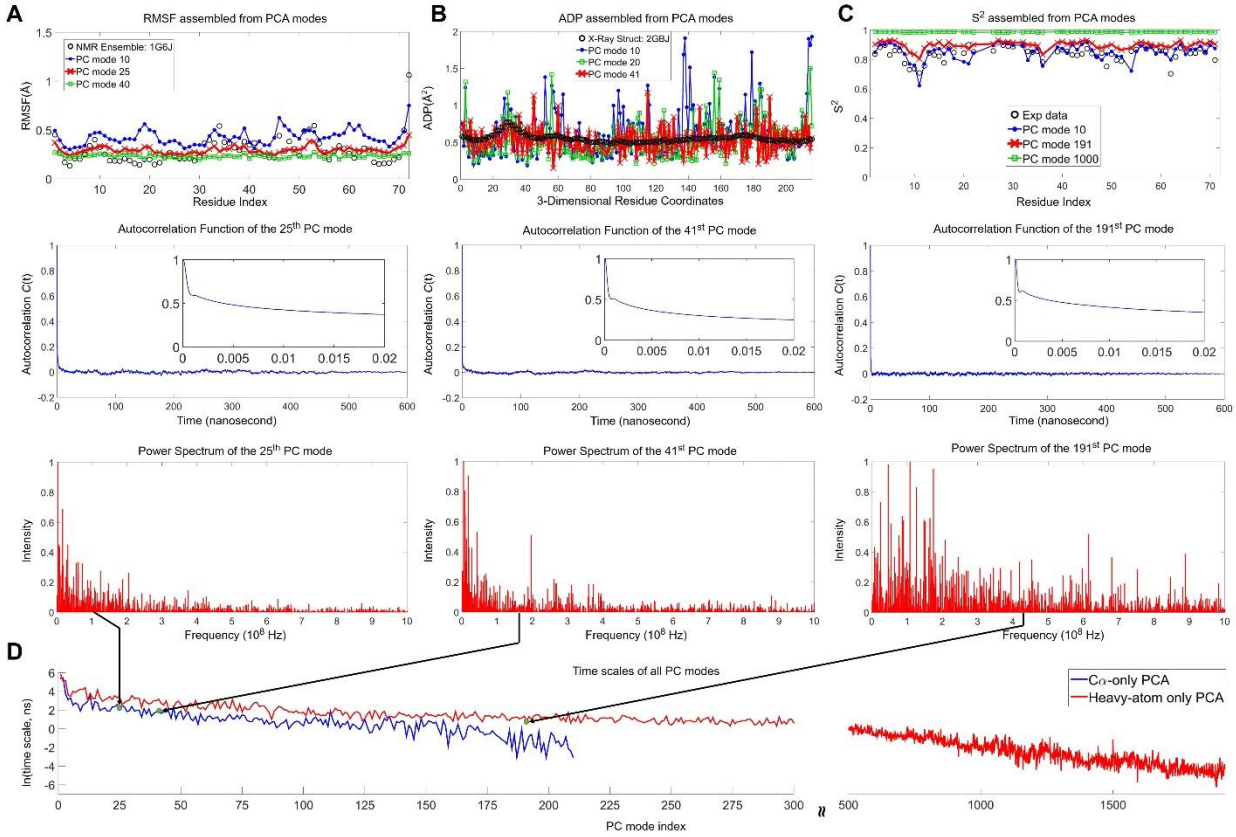


Figure 3. Time scale mapping for different experimental data that characterize protein dynamics. The first rows for **(A-C)** are, respectively, the oFP of RMSF derived from an NMR-determined ensemble (PDB: 1G6J), ADP values (only the diagonal elements are shown) and NMR-measured order parameters against 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 $<\Delta\mathbf{R}\Delta\mathbf{R}^T>_{MD,k}$ (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 **Equation 1**. 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 (Tjandra et al., 1995) 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.

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 (Best and Vendruscolo, 2004; Henry and Szabo, 1985; Yang et al., 2009)

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

(Equation 4)

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 S2C**), characterizes the time scale and size of motion for the observed order parameter profiles (Tjandra et al., 1995). 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)

(Lee et al., 1999) 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 (Nederveen and Bonvin, 2005).

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) (Atilgan et al., 2001; Li et al., 2017). 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 $\geq l$ (lower ANM modes have lower frequencies or smaller eigenvalues) was calculated as follows

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

(Equation 5)

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 $RMSF_{ANM,l}$ profile to identify the kth PC's $RMSF_{MD,k}$ 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 **STAR Methods** and **Figure 4A**.

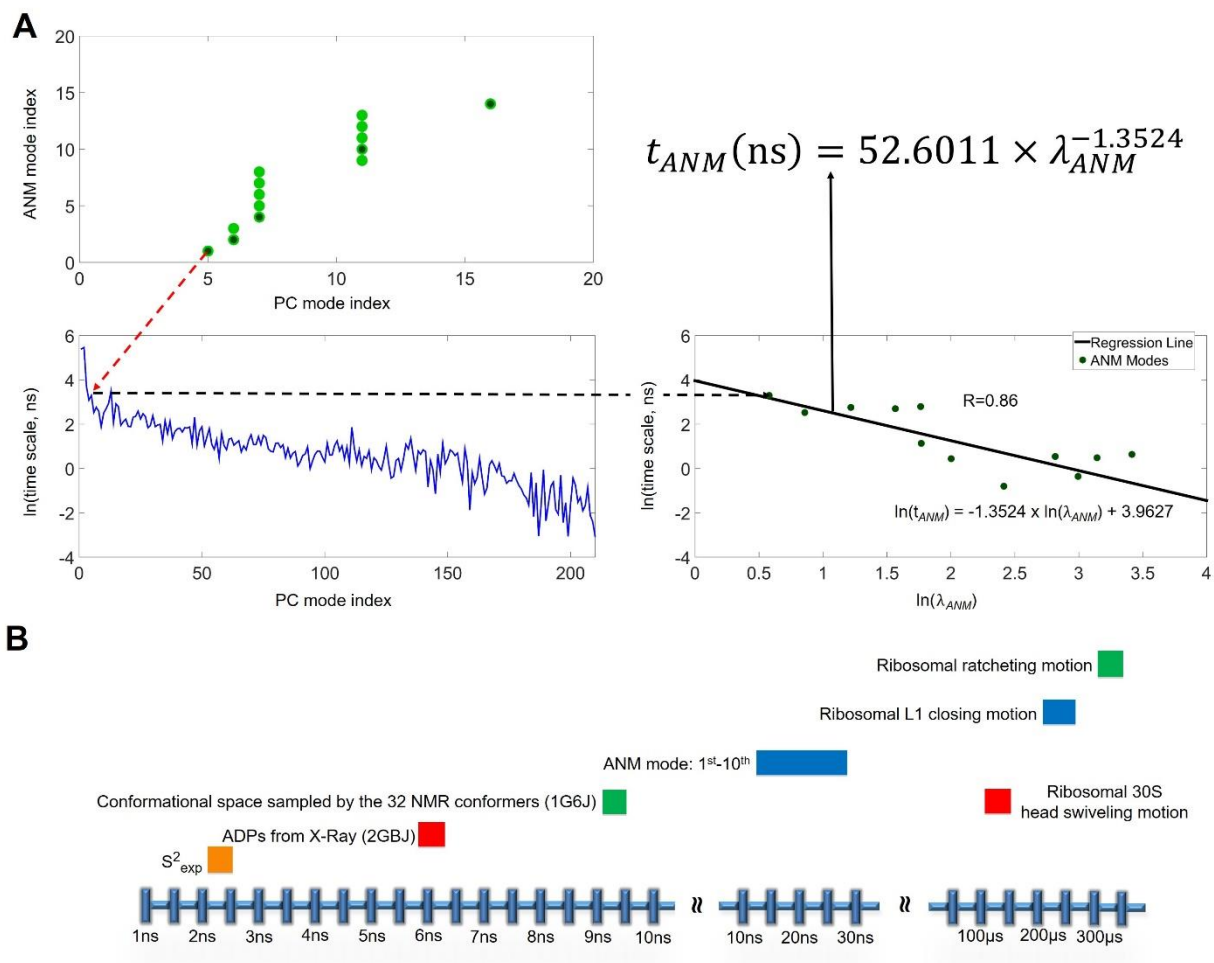


Figure 4. Frequency mapping for ANM to obtain the power law of τ_w as a function of ANM eigenvalues. (A) The ANM analysis and PCA of long MD trajectories are obtained for ubiquitin

(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 $\geq l$. The RMSF of each residue, $RMSF_{ANM,l}$, is the square root of the sum of x-, y- and z-components of its variance. This $RMSF_{ANM,l}$ 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 $t_{ANM}(\mathrm{ns}) = 52.6011 \times \lambda_{ANM}^{-1.3524}$ 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 $t_{ANM}(\mathrm{ns}) = 65.4392 \times \lambda_{ANM}^{-2.1386}$ 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, $t_{ANM}(\mathrm{ns}) = 52.6011 \times \lambda_{ANM}^{-1.3524}$.

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; also see **Figure S4B**) is nowhere final and presented as a seminal effort for further studies and improvements.

In the **Supplemental Information**, we provide a verification on how the time scales of several dynamics variables can be obtained using aforementioned general power law, as a function of ANM eigenvalue. Thus, a simple molecular timer that characterizes functional motions of biomolecules is obtained. **Figure 4B** summarizes the time scales of discussed experimental observables and ANM modes of interest.

Derive ANM power law for the sizes of functional motions

Since each ANM mode has been mapped to a PC mode, one can also 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 **STAR Methods**). 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 $t \sim \langle \sigma^2 \rangle$. Our derived $t^{1.33} \sim \langle \sigma^2 \rangle$ relationship lies closer to free diffusion than to a harmonic oscillator.

Analysis of ribosomal motions

The power laws are applied to analyze the ribosomal body rotation motion (ratcheting) (Tama et al., 2003; Wang et al., 2004) between the 30S and 50S subunits of the *Thermus thermophilus* ribosome (Tama et al., 2003), which is essential for ribosomal translocation during protein translation (Cornish et al., 2008; Frank and Agrawal, 2000). 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 the **STAR Methods** and **Supplemental Information**. We deform the ribosome by the size of mode 25 suggested by the variance power law and the rotated ribosome 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]) (Jenner et al., 2010; Tourigny et al., 2013).

The estimated timescale of the ribosomal body-rotation (ratcheting) according to simulation (Whitford et al., 2013) and stopped-flow (Cornish et al., 2008; Guo and Noller, 2012) 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 (Whitford et al., 2013). 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 ubiquitin and FGF2 only, which gives a $t_{ANM}(\mathrm{ns}) = 65.4392 \times \lambda_{ANM}^{-2.1386}$ (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 just 1.8-fold longer than 180 μs, the time scale estimated by the power law derived from all the

3 proteins ($t_{ANM}(\text{ns}) = 86.9387 \times \lambda_{ANM}^{-1.8886}$). 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 **Supplemental Information)**.

DISCUSSION

To summarize, the two power laws we derived are combined to give $\langle\sigma^2\rangle = 0.1223t^{1.3282}$. Such a "guided" (structure-encoded) diffusion motion ($t \sim \langle\sigma^2\rangle^{0.75}$) is slower than a harmonic oscillator that travels with a ($t \sim \langle\sigma^2\rangle^{0.5}$) (Atilgan et al., 2001) relation, but faster than the free diffusion ($t \sim \langle\sigma^2\rangle$) (McQuarrie, 2000). The predicted time scale using our power law at 100s of microseconds to milliseconds is estimated to have an error of 2-4 fold or less than an order of magnitude. We hope that the introduced method to turn ENMs into a molecular timer and sizer 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.

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).

AUTHOR CONTRIBUTIONS

L.W.Y. designed, developed and supervised the project. L.W.Y. and Y.J. implemented the Wiener–Khintchine theorem to obtain time-correlation functions. A.K. and L.W.Y. design the intensity-weighted time period, implemented by H.R.L. with zero paddings. K.T., L.W.Y. and J.C. jointly design the algorithms to compute the order parameters, implemented by H.R.L, K.T. and J.C., who also jointly generate the data for ADPs, RMSFs and (PC)mode-(ANM)mode mapping. K.C.C. performed the ANM analysis for ribosome. Y.Y.C. wrote part of the MATLAB codes for ANM analysis included in the bioStructureM package (https://github.com/Hong-Rui/bioStructureM). H.J.L., J.C. and L.W.Y. jointly drafted the first version of the paper, revised and finalized by all the co-authors.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

REFERENCES

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

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

Bahar, I., Lezon, T.R., Yang, L.-W., and Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. Annu. Rev. Biophys. *39*, 23–42.

Bakan, A., and Bahar, I. (2009). The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc. Natl. Acad. Sci. *106*, 14349–14354.

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

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

Case, D. a, Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J. Comput. Chem. *26*, 1668–1688.

Chandrasekaran, A., Chan, J., Lim, C., and Yang, L.-W. (2016). Protein dynamics and contact topology reveal protein-DNA binding orientation. J. Chem. Theory Comput. *12*, acs.jctc.6b00688.

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

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

Fei, J., Kosuri, P., MacDougall, D.D., and Gonzalez, R.L. (2008). Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. Mol. Cell *30*, 348–359.

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

Frank, J., and Agrawal, R.K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature *406*, 318–322.

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

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

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

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins Struct. Funct. Bioinforma. *65*, 712–725.

Huang, J., and Mackerell, A.D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. J. Comput. Chem. *34*, 2135–2145.

Jenner, L.B., Demeshkina, N., Yusupova, G., and Yusupov, M. (2010). Structural aspects of messenger RNA reading frame maintenance by the ribosome. Nat. Struct. Mol. Biol. *17*, 555–560.

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

Kitao, A., Hirata, F., and Gō, N. (1991). The Effects of Solvent on the Conformation and the Collective Motions of Protein: Normal Mode Analysis and Molecular Dynamics Simulations of Melittin in Water and in Vacuum. Chem. Phys. *158*, 447–472.

Kurkcuoglu, O., Doruker, P., Sen, T.Z., Kloczkowski, A., and Jernigan, R.L. (2008). The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. Phys. Biol. *5*, 046005.

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

Li, H., Sakuraba, S., Chandrasekaran, A., and Yang, L.-W. (2014). Molecular binding sites are located near the interface of intrinsic dynamics domains (IDDs). J. Chem. Inf. Model. *54*, 2275–2285.

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

Li, H., Chang, Y.-Y., Lee, J.Y., Bahar, I., and Yang, L.-W. (2017). DynOmics: dynamics of structural proteome and beyond. Nucleic Acids Res. *45*, W374–W380.

McQuarrie, D.A. (2000). The time-correlation function formalism II. In Statistical Mechanics, (University Science Books), pp. 543–592.

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

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

Okan, O.B., Atilgan, A.R., and Atilgan, C. (2009). Nanosecond motions in proteins impose bounds on the timescale distributions of local dynamics. Biophys. J. *97*, 2080–2088.

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

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

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

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. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature *468*, 713–716.

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

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. (2010). Atomic-Level Characterization of the Structural Dynamics of Proteins. Science (80-.). *330*, 341–346.

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

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

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

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

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

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

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

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

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

e1003003.

Yang, L.-W. (2011). Models with Energy Penalty on Interresidue Rotation Address Insufficiencies of Conventional Elastic Network Models. Biophys. J. *100*, 1784–1793.

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

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

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

Yang, L.-W., Kitao, A., Huang, B.-C., and Gō, N. (2014). Ligand-Induced Protein Responses and Mechanical Signal Propagation Described by Linear Response Theories. Biophys. J. *107*, 1415–1425.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contacts, Lee-Wei Yang (lwyang@life.nthu.edu.tw) or Akio Kitao (akitao@bio.titech.ac.jp).

METHOD DETAILS

MD Simulations

Detailed simulation protocols have been previously reported (Takemura and Kitao, 2012). Briefly, a 600ns MD simulation for Ubiquitin was performed using the PMEMD module of AMBER 10 (Case et al., 2005) with the ff99SB force field (Hornak et al., 2006), while CHARMM36 (Huang and Mackerell, 2013) force field was used to perform 200ns MD simulations for FGF2 and HPNAP with NAMD 2.10 (Phillips et al., 2005). 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.

Principal component analysis (PCA) of the covariance matrix computed from MD trajectories

Proteins in the snapshots were iteratively superimposed onto their mean structure, $<\mathbf{R}>$, and the covariance matrix, $<\Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}}>$, is constructed using a deviation matrix $\mathbf{Q}$ and decomposed into its corresponding eigenvalues and eigenvectors such that:

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

(**Equation S1**)

$\Delta\mathbf{R}$ is the $3N$-dimensional deviation vector, and $\mathbf{Q}$ 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 $\mathbf{Q}$ 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 of a trajectory are then projected onto the principal components to form a projection matrix $\mathbf{U}_{3N\times M}$ as

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

(**Equation S2**)

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.

Wiener–Khintchine theorem (WKT)

In this section, we show the detailed derivations and implementations of the autocorrelation functions from the PC modes, obtained from MD simulations, using the Wiener–Khintchine theorem. For the kth mode and a trajectory containing M snapshots, we can express the trajectory as a discrete function of time s as $u_k(s) = \{u_{k0}, u_{k1}, \ldots u_{kM-1}\}$. Our goal here is to obtain the autocorrelation for the 'particle motion' (the projections of the MD snapshots) at the nth time point. In statistical mechanics, it is, by definition, the ensemble average of the correlation of particle positions at a time shift $t \equiv n\Delta t$, where Δt is the time step. For long-enough equilibrium simulation, the ensemble average can be replaced by the time average under the ergodic hypothesis such that

$$\begin{aligned} C_k(t) &\equiv < u_k(s)u_k(t+s) > \\ &= < u_k(s)u_k(n\Delta t + s) > \\ &\approx \frac{1}{M-n}\sum_{i=1}^{M-n} u_k(s_i)u_k(s_i + n\Delta t) \end{aligned}$$

(**Equation S3**)

where s_i is the time point when we start to calculate the time-correlation between the u_k at s_i and the u_k after n time-steps; n goes from 0 to $(M - 1)$. The summation runs from 1 to $(M - n)$ time point since we are only interested in a n-time-step shift. The computation time of such a calculation grows with $O(M^2)$ for all the possible t to be obtained. Long computation time is inevitable especially for a long trajectory, repeated for many modes ($3N - 6$, where N is the degrees of freedom being analyzed). For our purpose here as well lessening the computational burden, we want to obtain the time-correlation function by first Fourier transforming snapshot projections on a given mode into their counterparts in the frequency domain, and then inverse transform the power spectrum back to obtain the $C_k(t)$. This mathematical treatment is known as Wiener–Khintchine theorem (McQuarrie, 2000) in statistical mechanics, which proves that the autocorrelation function is equal to the inverse Fourier transformation of the power spectrum of the trajectories such that

$$C_k(t) = \frac{1}{2M}\sum_{f=0}^{2M-1} S_k(f)\, e^{\frac{2\pi i f t}{2M}}$$

(**Equation S4**)

where power spectrum $S_k(\omega)$ is defined as the magnitude of the Fourier transformation of snapshot projections on the kth PC mode $\widetilde{u_k}(\omega) = \{\widetilde{u_{k0}}, \widetilde{u_{k1}}, \ldots \widetilde{u_{k2M-1}}\}$, such that $S_k(\omega) = \{|\widetilde{u_{k0}}|^2, |\widetilde{u_{k1}}|^2, \ldots |\widetilde{u_{k2M-1}}|^2\}$. There are $2M$ components in $u_k(s)$ instead of M components after we pad M zeros after the original M u_k projections such that $u_k(s) = \{u_{k0}, u_{k1}, \ldots u_{kM-1}, 0^1, 0^2, \ldots 0^M\}$. This is to make sure that we obtain M positive frequency values.

In addition, one can notice that in the last line of **Equation S3**, termed as brute force method herein, for the nth value of correlation function, the summation goes from 1 to ($M - n$), leading to less addends in the summation as n increases. Therefore, we need to pad enough zeros at the end of the time series as the input of DFT in order to make sure the Fourier transformation reproduces correct results. In our case, we pad M zeros in the end of the sequence so that after the inverse transformation we get back a sequence of length $2M$. For our needs, we only take the sequence in the positive time domain, with M time points. The comparison between Wiener–Khintchine theorem and the brute force method shows that Wiener–Khintchine theorem can indeed reproduce correct time correlation functions, evident that our method is applicable in computing autocorrelation functions with a time complexity of $O(M \ln M)$.

Intensity-Weighted Period (IWP), Relaxation Time and Characteristic Time

The IWP for a specific PC mode k is computed as

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

(**Equation S5**)

where $S_k(\omega_i)$ is the power for the angular frequency, ω_i.

While, the relaxation of $C_k(t)$ is modeled as an exponential function

$$C_k(t) = A \exp\left(-\frac{t}{\tau_r}\right)$$

(**Equation S6**)

The values of $C_k(t)$ from unity (when $t = 0$) to the first instance that the function reaches zero ($<10^{-5}$) were used to fit the exponential and obtain τ_r.

The characteristic time is defined as

$$\tau_c = \frac{\sum_{i=1}^{M} C_k(t_i)\Delta t}{C_k(0)}$$

(**Equation S7**)

where $\Delta t = 0.1$ ps is the time interval between two consecutive snapshots.

Fluctuation Profile Matching (FPM)

The FPM method is used to identify a subset of PC modes that correspond to dynamical profiles including RMSF, ADP or order parameters. This allows the assignment of timescale and variance of the slowest PC mode to the profile.

FPM takes 2 profiles of the same dynamical variable (one experiment and one theoretical, or two theoretical ones) and assesses their degree of agreement via Pearson's correlation coefficient (Pearson, 1895), which ranges from -1 to 1, while gradually removing some constituent modes.

FPM is used in the following sections to estimate the timescale of motions producing the RMSF, ADP and order parameter profiles of ubiquitin from the IWP of PC modes. In a later section, FPM is used to assign the timescale and variance to each ANM mode which provides the data for fitting the power laws.

RMSF and ADP Profiles for Ubiquitin derived from X-ray/NMR-determined Structures and MD
The $RMSF_{exp}$ profile was calculated by treating the ensemble of 32 NMR-determined conformers (PDB ID: 1G6J) of ubiquitin (Cα atoms only) as snapshots and obtaining the $<\Delta\mathbf{R}\Delta\mathbf{R}^{T}>_{exp,1G6J}$ matrix by **Equation S1**. Then the RMSF of each residue is the square-root of the sum of the x-, y- and z-dimension variances for the respective residues taken from the diagonal of $<\Delta\mathbf{R}\Delta\mathbf{R}^{T}>_{exp,1G6J}$. The $RMSF_{MD,k}$ profile is computed using the same method but with their respective $<\Delta\mathbf{R}\Delta\mathbf{R}^{T}>_{MD,k}$ matrix in **Equation S8**.

To obtain ADP_{exp}, the Cα atoms of the ubiquitin X-ray structure PDB ID: 2GBJ (chain B) is superimposed onto the mean structure of MD snapshots. The rotation matrix $\mathbf{R'}_{3\times3}$ used to remove rotational differences between the 2 structures is obtained. Then, the ADP matrix, $\mathbf{U'}_{3\times3}$, for each residue is rotated to obtain the superimposed ADP matrices ($\mathbf{R'U'R'}^{T}$) (Eyal et al., 2007). From the ADP matrix, the ADP dynamical variable of each residue consisting of 3 variance components and 3 unique covariance components (the xy- and yx- components are the same and so are xz- and zx- components, etc.) are extracted to form the 6*N*-dimensional ADP_{exp} profile.

While, the ADP matrix from MD for each Cα atom is the atom's x-, y- and z-dimension covariance matrix extracted from the full $<\Delta\mathbf{R}\Delta\mathbf{R}^{T}>_{MD,k}$ matrix. The component of this matrix was extracted exactly as above to form the $ADP_{MD,k}$ profile.

An example of assigning the time-scale of an observed dynamical variable from the IWP of an FPM matched PC mode
For instance, we can take an ensemble of 32 NMR-determined conformers (PDB ID: 1G6J) and form a Cα-only $3N\times3N$ $<\Delta\mathbf{R}\Delta\mathbf{R}^{T}>_{exp,1G6J}$ matrix to obtain $RMSF_{exp}$ (**Figure 3A, top**), see above for details. There are 210 ($3N - 6 = 210$) $RMSF_{MD,k}$ profiles, each calculated using all the PC modes $\geq k^{th}$ mode. Among the 210 RMSF profiles, $RMSF_{MD,k=25}$ 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 $\tau_{w,k=25}$ = 9.32 ns (**Figure 3A, middle and bottom**).

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**.

Theoretical Covariance Matrix Comprising a set of PC modes Derived from MD simulations
The eigenvectors and eigenvalues are obtained as per **Equation S1** with the first *k*-1 modes (lower PC modes have larger variances or eigenvalues) being removed such that

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

(**Equation S8**)

where $\mathbf{V}_a$ is the eigenvector with its corresponding eigenvalue, λ_a, taken from the diagonal of $\mathbf{\Lambda}$ in **Equation S1** for the *a*'th PC mode.

Take for example, to compute the $<\Delta\mathbf{R}\Delta\mathbf{R}^{\boldsymbol{T}}>_{MD,5}$ covariance matrix, the four PCs with the largest eigenvalues/variances are not included in the summation, thus excluding their contribution to the covariance matrix.

Order Parameter Profiles derived from NMR-determined structural ensemble and MD snapshots

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 (Best and Vendruscolo, 2004; Henry and Szabo, 1985; Yang et al., 2009)

$$S_{ij}^2 = \frac{3}{2}\left(< x_{ij}^2 >^2 + < y_{ij}^2 >^2 + < z_{ij}^2 >^2 + 2 < x_{ij}y_{ij} >^2 + 2 < x_{ij}z_{ij} >^2 + 2 < y_{ij}z_{ij} >^2\right) - \frac{1}{2}$$

(**Equation S9**)

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.

To obtain the MD-derived $S^2_{MD,k}$ profile for all the PC modes $\geq k^{th}$ mode, the protein structure in each snapshot is rebuilt using all the PC modes $\geq k^{th}$ mode which is then used to obtain the -NH bond vectors for calculating $S^2_{MD,k,ij}$ with **Equation S9**.

The method used to rebuild the structures as a function of constituent modes is as follows, for ubiquitin's structure in snapshot *m*, the coordinates of its atoms ($\mathbf{R}_m$) can be expressed as

$$\mathbf{R}_m = < \mathbf{R} > + \sum_{a=1}^{3N-6} \left((\mathbf{R}_m - < \mathbf{R} >) \cdot \mathbf{V}_a\right)\mathbf{V}_a$$

(**Equation S10**)

Then, the atom coordinates of snapshot *m* rebuilt using all the PC modes $\geq k^{th}$ mode can be obtained by the following equation

$$\mathbf{R}_{m,k} = < \mathbf{R} > + \sum_{a=k}^{3N-6} \left((\mathbf{R}_m - < \mathbf{R} >) \cdot \mathbf{V}_a\right)\mathbf{V}_a$$

(**Equation S11**)

where <**R**> represents the mean coordinate of the structure over M snapshots, and $\mathbf{V}_a$ is the ath PC eigenvector.

Now, for each snapshot $\mathbf{R}_{m,k}$, we compute the normalized NH bond vector r_{ij} for every residue (except for prolines). The averages of x_{ij}^2, y_{ij}^2, z_{ij}^2, $x_{ij}y_{ij}$, $x_{ij}z_{ij}$ and $y_{ij}z_{ij}$ over the whole trajectory were calculated and used to obtain $S^2_{MD,k,ij}$.

The experimental order parameters, S^2_{exp}, for ubiquitin were taken from Tjandra et al (Tjandra et al., 1995).

Derivation of General Elastic Network Model (ENM) – the relation between covariance and Hessian

In the equilibrium state, the protein system is harmonically approximated and assumed to have the Hamiltonian

$$H_0 = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + \frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}$$

(**Equation S12**)

Where **p** is the $3N\times1$ momentum vector in the Cartesian coordinate system, and $\mathbf{M}^{-1}$ is the inverse of a $3N\times3N$ diagonal mass matrix comprising elements of triplicate mass of every node (Yang, 2011; Yang and Chng, 2008; Yang et al., 2014). The superscript T denotes matrix transposition. $\Delta\mathbf{R}$ is a $3N\times1$ displacement vector and **H** is the symmetric Hessian matrix, or, the force constant tensor of the system, and have the meaning of coupling force constants in each pairwise connected degrees of freedom.

Here we note that since **H** is diagonalizable, for the ease of the following derivations of the ensemble average of positional covariance, we transform **H** in normal-mode space, such that

$$\begin{aligned}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R} &= \Delta\mathbf{R}^T(\mathbf{U}\mathbf{D}\mathbf{U}^T)\Delta\mathbf{R}\\ &= (\mathbf{U}^T\Delta\mathbf{R})^T\mathbf{D}(\mathbf{U}^T\Delta\mathbf{R})\\ &= \mathbf{q}^T\mathbf{D}\mathbf{q}\end{aligned}$$

(**Equation S13**)

Where **D** is the diagonal eigenvalue matrix, **U** is the unitary eigenvector matrix and **q** is the displacement vector in normal-mode space with the relationship $\mathbf{q} = \mathbf{U}^T\Delta\mathbf{R}$.
Since unitary transformation conserves length so that we can change the integration to normal-mode space, which is readily solvable.

Now, from **eqs S12 and S13**, we write H_0 in the expanded form as

$$H_0 = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + \frac{1}{2}\mathbf{q}^T\mathbf{D}\mathbf{q} = \frac{1}{2}\sum_{k=1}^{3N}\frac{p_k^2}{m_k} + \frac{1}{2}\sum_{k'=1}^{3N}\lambda_{k'}q_{k'}^2$$

(Equation S14)

In the above expression, k and k' represent the index through all $3N$ degrees of freedom, and m_k is the mass elements in **M** with N triplicates.

In canonical ensemble we write down the partition function with phase space formulation and solve it by expanding all the terms

$$\begin{aligned}
Z_0 &= \iint e^{-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + \frac{1}{2}\mathbf{q}^T\mathbf{D}\mathbf{q}\right)}\, d^{3N}\mathbf{p}\, d^{3N}\mathbf{q} \\
&= \int e^{-\frac{\beta}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}}\, d^{3N}\mathbf{p} \int e^{-\frac{\beta}{2}\mathbf{q}^T\mathbf{D}\mathbf{q}}\, d^{3N}\mathbf{q} \\
&= \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{-\frac{\beta}{2}\sum_{k=1}^{3N}\frac{p_k^2}{m_k}}\, dp_1 dp_2 \dots dp_{3N} \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} e^{-\frac{\beta}{2}\sum_{k'=1}^{3N}\lambda_{k'}q_{k'}^2}\, dq_1 dq_2 \dots dq_{3N} \\
&= \prod_{n=1}^{N}\left[\left(\int_{-\infty}^{\infty} e^{-\frac{\beta p_n^2}{2m_n}}\, dp_n\right)^3\right]\prod_{k'=1}^{3N}\left[\int_{-\infty}^{\infty} e^{-\frac{\beta\lambda_{k'}q_{k'}^2}{2}}\, dq_{k'}\right] \\
&= \prod_{n=1}^{N}\left[\left(\frac{2\pi m_n}{\beta}\right)^{\frac{3}{2}}\right]\sqrt{\frac{(2\pi)^{3N}}{\beta^{3N}(\lambda_1\lambda_2\dots\lambda_{3N})}} \\
&= \left(\frac{2\pi}{\beta}\right)^{\frac{3N}{2}}(m_1 m_2 \dots m_N)^{\frac{3}{2}}\sqrt{\frac{(2\pi)^{3N}}{\beta^{3N}\det\mathbf{H}}}
\end{aligned}$$

(Equation S15)

where $\beta=1/k_BT$; k_B is the Boltzmann constant and T is the absolute temperature.
In the equation above we use the property that for a symmetric matrix the product of its eigenvalues is the determinant of itself.

To calculate the positional covariance, we note one derivative

$$\begin{aligned}
\frac{\partial(\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R})}{\partial H_{ij}} &= \frac{\partial[\sum_{k=1}^{3N}\Delta R_k(\sum_{l=1}^{3N}H_{kl}\Delta R_l)]}{\partial H_{ij}} \\
&= \Delta R_i\frac{\partial(\sum_{l=1}^{3N}H_{il}\Delta R_l)}{\partial H_{ij}} \\
&= \Delta R_i\frac{\partial H_{ij}}{\partial H_{ij}}\Delta R_j \\
&= \Delta R_i\Delta R_j \\
&= (\Delta\mathbf{R}\Delta\mathbf{R}^T)_{ij}
\end{aligned}$$

(Equation S16)

Now

$$\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \rangle = \frac{\iint \Delta\mathbf{R}\Delta\mathbf{R}^T e^{-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}+\frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}\right)}\, d^{3N}\mathbf{p}\, d^{3N}\Delta\mathbf{R}}{Z_0}$$
$$= \frac{\iint \Delta\mathbf{R}\Delta\mathbf{R}^T e^{-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}+\frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}\right)}\, d^{3N}\mathbf{p}\, d^{3N}\Delta\mathbf{R}}{\iint e^{-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}+\frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}\right)}\, d^{3N}\mathbf{p}\, d^{3N}\Delta\mathbf{R}}$$

(Equation S17)

For each element in the covariance matrix

$$\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \rangle_{ij} = \frac{\iint (\Delta\mathbf{R}\Delta\mathbf{R}^T)_{ij} e^{-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}+\frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}\right)}\, d^{3N}\mathbf{p}\, d^{3N}\Delta\mathbf{R}}{\iint e^{-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}+\frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}\right)}\, d^{3N}\mathbf{p}\, d^{3N}\Delta\mathbf{R}}$$
$$= -\frac{2}{\beta}\frac{\partial}{\partial H_{ij}} \ln Z_0$$
$$= -\frac{2}{\beta}\frac{\partial}{\partial H_{ij}}\left(\frac{3N}{2}\ln\frac{2\pi}{\beta} + \frac{3}{2}\ln(m_1 m_2 \dots m_N) + \frac{3N}{2}\ln\frac{2\pi}{\beta} - \frac{1}{2}\ln\det\mathbf{H}\right)$$
$$= \frac{1}{\beta}\frac{1}{\det\mathbf{H}}\frac{\partial\det\mathbf{H}}{\partial H_{ij}}$$
$$= \frac{1}{\beta}\frac{1}{\det\mathbf{H}}\frac{\partial\left(\sum_{k=1}^{3N} H_{ik}C_{ik}\right)}{\partial H_{ij}}$$
$$= \frac{1}{\beta}\frac{1}{\det\mathbf{H}}\sum_{k=1}^{3N}\delta_{jk}C_{ik}$$
$$= \frac{1}{\beta}\frac{1}{\det\mathbf{H}}C_{ij}$$
$$= \frac{1}{\beta}\frac{1}{\det\mathbf{H}}[(\mathrm{adj}\mathbf{H})^T]_{ij}$$
$$= \frac{1}{\beta}\frac{1}{\det\mathbf{H}}\left[\det\mathbf{H}\cdot(\mathbf{H}^{-1})^T_{ij}\right]$$
$$= \frac{1}{\beta}(\mathbf{H}^{-1})_{ij}$$

(Equation S18)

Where C_{ik} in the above derivation is the *ik*-element of the cofactor matrix of the Hessian. Here we use Laplace formula to expand the Hessian determinant and using the properties that (1) the adjugate matrix is the transpose of cofactor matrix, (2) inverse of any nonsingular matrix **A** is $\mathbf{A}^{-1} = \frac{1}{\det\mathbf{A}}\mathrm{adj}\mathbf{A}$ and (3) the inverse of any symmetric matrix, in our case Hessian, is still symmetric.

Finally, in matrix form we write

$$\langle \Delta\mathbf{R}\Delta\mathbf{R}^T \rangle = \frac{1}{\beta}\mathbf{H}^{-1}$$

(**Equation S19**)

which is the expected result by analogy to one-dimensional harmonic oscillator system.

This result holds true for all types of ENMs as long as the system is harmonically approximated. For ENMs that have translational and/or rotational invariance (Yang, 2011), which makes the rank of **H** less than its dimension, $\mathbf{H}^{-1}$ is solved by eigenvalue decomposition followed by removing the terms containing zero eigenvalues.

The Anisotropic Network Model (ANM)

The Anisotropic Network Model (ANM) is a type of ENM with the following potential energy function

$$U_{ANM} = \sum_{i.j=1}^{N} \frac{\gamma}{2}\left(L_{ij} - L_{ij}^0\right)^2 \theta\left(r_c - L_{ij}^0\right) = \frac{1}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R}$$

(**Equation S20**)

where the index *i* and *j* represents the nodes (Cα atoms) that runs from 1 to *N*. The γ represents the universal spring constant of the system. L_{ij} is the current distance between the ith and j^{th} nodes while L_{ij}^0 is the equilibrium distance between *i* and *j*. When L_{ij}^0 is $\leq r_c$, the cut-off distance, then these two nodes are treated as connected. This is expressed as a Heaviside-step function θ in **Equation S20**.

All the elements of the Hessian matrix **H** can be calculated as the second derivative of the potential energy function. Details of this can be found in the supplementary information of our previous work (Yang, 2011). The eigenvector ($\boldsymbol{V}_{ANM}$) and eigenvalue (λ_{ANM}) of each ANM mode is calculated from the eigen-decomposition of **H such that**

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

(**Equation S21**)

ANM analysis can be performed on the DynOmics Portal 1.0 (Li et al., 2017): http://dyn.life.nthu.edu.tw/oENM/.

Theoretical Covariance Matrix, RMSF and ADP from ANM With Removed Modes

The ANM derived covariance matrix for all ANM modes ≥ *l* (lower ANM modes have lower frequencies or smaller eigenvalues) can be calculated as follows

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

(**Equation S22**)

where $\boldsymbol{V}_a$ and λ_a are the eigenvector and eigenvalue of the a'th ANM mode.

The $RMSF_{ANM,l}$ and $ADP_{ANM,l}$ 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.

ANM Mode Time Scale Assignment Through ANM-PCA Mode Mapping

Every ANM mode l was mapped to a PC mode k* from MD by comparing the $RMSF_{ANM,l}$ with the $RMSF_{MD,k}$ profiles for different k's using FPM to identify the $k = k$* with the highest correlation. Every ANM mode l was then assigned the IWP ($\tau_{w,k*}$) of the mapped PC mode k*.

The ANM-PCA mode mapping might result in a PC mode getting mapped to multiple ANM modes. The ANM mode with the highest correlation with the PC mode was kept while the rest were removed. This filtered set of ANM-PCA mode mapping was used to fit the power laws.

Fitting the Time Power Laws

The relationship between an ANM mode's eigenvalue (λ_{ANM}) and mapped time scale (t_{ANM}) was modeled as a power law in the form of $t_{ANM} = c \times \lambda_{ANM}^{d}$, with the constants c and d to be found through the fitting procedure described below.

The natural log of the power law, $\ln(t_{ANM}) = \ln(c) + d \ln(\lambda_{ANM})$, was used to get a linear function of $\ln(t_{ANM})$ with respect to $\ln(\lambda_{ANM})$. Least squares fitting was then used to obtain the parameters c and d.

Fitting the Variance Power Law

Given the mapping between ANM modes and PCA modes, the λ_{ANM} and λ_{PCA} (PC mode's eigenvalue/variance) were used to fit a variance power law such that $\lambda_{PCA} = a \times \lambda_{ANM}^{b}$, where a and b are constants. The constants were found by following the same procedure used to fit the time power laws.

Identifying the Functional Modes Corresponding to Ribosomal Body Rotation (Ratcheting) and Head Swiveling Motions

The body rotation is known to be generally around an axis that connects the mass centers of the 30S and 50S subunits of the ribosome (Kurkcuoglu et al., 2008; Tama et al., 2003; Wang et al., 2004). Thus, the axis of rotation, $\mathbf{r}_B$ the vector connecting the COMs of the 2 subunits as observed in the non-rotated x-ray crystallographically resolved conformation (PDB ID: 4V6F). While for the head swiveling motion's axis of rotation, $\mathbf{r}_H$, it is the vector between the COMs of the 30S body and 30S head. Where the 30S head is defined as the residues 921-1396 of the 16S rRNA

complexed with the ribosomal proteins S3, S7, S9, S10, S13, S14, S19 and Thx. While, the 30S body is defined as residues 5-920 and 1399-1543 of the 16S rRNA complexed with ribosomal proteins S2, S4, S5, S6, S8, S11, S12, S15, S16, S17, S18 and S20.

ANM analysis was performed on the non-rotated x-ray solved conformation. The kth ANM eigenvector ($\mathbf{V}_k$) was used to deform the non-rotated conformation by

$$\mathbf{R}_k = \mathbf{R}_0 \pm \sqrt{k_B T/\gamma}\mathbf{V}_k$$

(**Equation S23**)

where $\mathbf{R}_0$ is a $3N$ vector containing the coordinates of the N atoms in the non-rotated conformation and $\mathbf{R}_k$ is the conformation deformed from $\mathbf{R}_0$ along the kth ANM mode by setting $k_B T/\gamma$ to be unity. For the purpose of obtaining axis of rotation, the results do not change as long as $k_B T/\gamma$ is set to be a small value.

The ratcheting motion is relative to the 50s subunit and therefore the deformed conformation was first superimposed at the 50S subunit. To capture the head swiveling motion, we superimposed the deformed conformation at the 50S subunit and 30S' body part to avoid capturing the ratcheting motion and examined the rotation of the 30S head.

Considering the transition from non-rotated ($\mathbf{R_0}$) to deformed conformation ($\mathbf{R_k}$) takes place at a time interval Δt (in seconds), the angular momentum vector ($\mathbf{L}_k$) for the kth mode is the cross product between the mass-weighted position of atoms and atom velocity, such that

$$\mathbf{L}_k = \sum_i m_i \mathbf{R}_{0,i} \times \mathbf{v}_{k,i}$$

(**Equation S24**)

where $\mathbf{R}_{0,i}$ is the ith atom's position vector in the non-rotated conformation, m_i is the atom mass of i and the velocity vector $\mathbf{v}_{k,i} = (\mathbf{R}_{k,i} - \mathbf{R}_{0,i})/\Delta t$ where $\mathbf{R}_{k,i}$ is the position vector of atom i in the conformation deformed along the kth mode.

$\mathbf{L}_k$ is related to the angular velocity vector ($\boldsymbol{\omega}_k$) by

$$\begin{aligned} \mathbf{L}_k &= \mathbf{I}\boldsymbol{\omega}_k \\ \boldsymbol{\omega}_k &= \mathbf{I}^{-1}\mathbf{L}_k \\ \boldsymbol{\omega}_k \Delta t &= \mathbf{I}^{-1}\mathbf{L}_k \Delta t = \mathbf{I}^{-1} \sum_i m_i \mathbf{R}_{0,i} \times \left(\mathbf{R}_{k,i} - \mathbf{R}_{0,i}\right) \end{aligned}$$

(**Equation S25**)

The direction of $\boldsymbol{\omega}_k$ is the *axis of rotation* for mode k while $|\boldsymbol{\omega}_k|$ is the angular velocity in the unit of radians/second and therefore $|\boldsymbol{\omega}_k|\Delta t$ gives us the angle of rotation in radians (the length of Δt does not change the result).

The deviation angle θ between $\mathbf{r}_{B\ or\ H}$ and rotation axis $\boldsymbol{\omega}_k$ from mode k were computed for the slowest 50 ANM modes, where $\cos\theta = (\mathbf{r}_{B\ or\ H} \cdot \boldsymbol{\omega}_k)/(\|\mathbf{r}_{B\ or\ H}\|\|\boldsymbol{\omega}_k\|)$. The mode with the smallest angle deviating from $\mathbf{r}_B$ (or $\mathbf{r}_H$) was chosen as the functional mode (k^f) for the ratcheting (or for swiveling) motion.

Among the slowest 50 ANM modes, the *rotation axis* derived from mode 25 was identified to have the smallest deviation angle ($\theta = 11.6^o$) from the vector that connects the mass centers of 30S and 50S subunits. Therefore, mode 25 represents the ribosomal ratcheting motion. As for the head swiveling motion, ANM mode 32 has the smallest angle (θ = 3.9°) deviating from the axis of rotation.

Predicting the Conformation of the Ribosome After the Ratcheting and Head Swiveling Motions

After identifying the ANM mode (k^f) that corresponds to the body rotation/head swiveling motion, the conformation of the ribosome could be obtained by

$$\mathbf{R}_{k^f} = \mathbf{R}_0 \pm \sqrt{\sigma_{k^f}{}^2}\mathbf{V}_{k^f}$$

(**Equation S26**)

where variance ($\sigma_{k^f}^2$) of mode k^f can be predicted by the variance power law $\sigma_{k^f}^2 = 46.0538\ \lambda_{k^f}^{-2.5085}$ with λ_{k^f}, the eigenvalue of mode k^f.

Identifying the Functional Modes and Conformations Corresponding to the L1 Stalk Closing Motion

The L1 stalk is defined as the helices 76-78 (residues 2093-2196) of the 23S rRNA complexed with L1 protein and the COM of E-site is defined by the COM of E-site tRNA as observed in the non-rotated ribosome (PDB ID: 4V6F) (Fei et al., 2008).

To identify the L1 stalk's functional mode and conformation, the non-rotated ribosome (PDB ID: 4V6F) was first deformed along all 50 modes using the variance power law as in **Equation S26**. The deformed conformations were then superimposed at the 50S subunit as described above. Finally, we identify the mode with the shortest distance between the COM of the deformed L1 stalk and the E-site in the non-rotated ribosome.

The conformations deformed with ANM mode 27 and 28 have the two closest distance (~26 Å and ~30 Å, respectively) between the COM of L1 stalk and the COM of the E-site in the ribosome. We treat both modes as part of the L1 stalk closing motion.

QUANTIFICATION AND STATISTICAL ANALYSIS

To evaluate the quality of the fit between the power laws and timescale/variance, the Pearson correlation coefficient (Pearson, 1895) between the predicted timescale/variance with the MD derived timescale/variance was compared.

The Person correlation coefficient is also used in FPM to evaluate the goodness of fit between two profiles of dynamics variables.

DATA AND SOFTWARE AVAILABILITY

The input files used to perform the MD simulations are included as zipped files as part of the Supplementary Information.

ANM, WKT, IWP, FPM and the power laws are implemented in Matlab and the codes are available at https://github.com/Hong-Rui/bioStructureM. The MD trajectories of Ubiquitin, FGF2 and HPNAP can be requested by e-mail.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Matlab R2017a	MathWorks	https://www.mathworks.com/
AMBER 10	(Case et al., 2005)	http://ambermd.org/
NAMD 2.10	(Phillips et al., 2005)	http://www.ks.uiuc.edu/Research/namd/
Wiener–Khintchine theorem (WKT)	(McQuarrie, 2000)	https://github.com/Hong-Rui/bioStructureM
Intensity Weighted Period (IWP)	This paper	https://github.com/Hong-Rui/bioStructureM
Fluctuation Profile Matching (FPM)	This paper	https://github.com/Hong-Rui/bioStructureM
Power Laws	This paper	https://github.com/Hong-Rui/bioStructureM