bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

An efficient timer and sizer of protein motions reveals the time scales of functional dynamics in structured biomacromolecules

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;

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The clock of life ticks as fast as how efficiently proteins could perform their functional dynamics. Protein complexes execute functions via several large-scale intrinsic motions across multiple conformational states, which occur at a timescale of nano- to milliseconds for well-folded proteins. Computationally expensive molecular dynamics (MD) simulation has been the only theoretical tool to time and size these motions, though barely to their slowest ends. Here, we convert a simple elastic network model (ENM), which takes a few seconds (ubiquitin) to hours (ribosome) for the analysis, into a molecular timer and sizer to gauge the slowest functional motions of proteins. 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 $t(\text{ns}) = 86.9\lambda_{ENM}^{-1.9}$ and $\sigma^2(\text{Å}^2) = 46.1\lambda_{ENM}^{-2.5}$ are established allowing the characterization of the time scales of Nuclear Magnetic Resonance (NMR)-solved conformers, crystallographic anisotropic displacement parameters, and important ribosomal motions, as well as motional sizes of the latter.

Author Summary

The time scale of biological processes is governed by protein functional dynamics that often corresponds to the largest conformational spread and the longest time scales among all possible motions. Current simulation methodologies cannot reach the slowest, often functional, motions especially for supramolecular machineries. Borrowing the spring-bead model used in polymer physics since 60s, the efficient elastic network model (ENM), introduced in 90s, captured

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

all modes of protein motions but largely underestimated the time scales of slowest modes due to its harmonic approximation.

Here we map water-damped modes sampled by MD simulations to corresponding ones in ENM and thereby establish 2 power laws that describe the authentic time scales and sizes for the slowest anharmonic modes as functions of ENM eigenvalues. With that, we adequately describe the sizes and time scales for three proteins, confirmed by NMR spectroscopy, and ribosome that contains ~0.2 million heavy atoms (~ 20 thousands coarse-grained nodes).

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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 **Fig 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 **Fig 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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

accurate methods are needed to predict the "time periods" and absolute variances of these anharmonic modes.

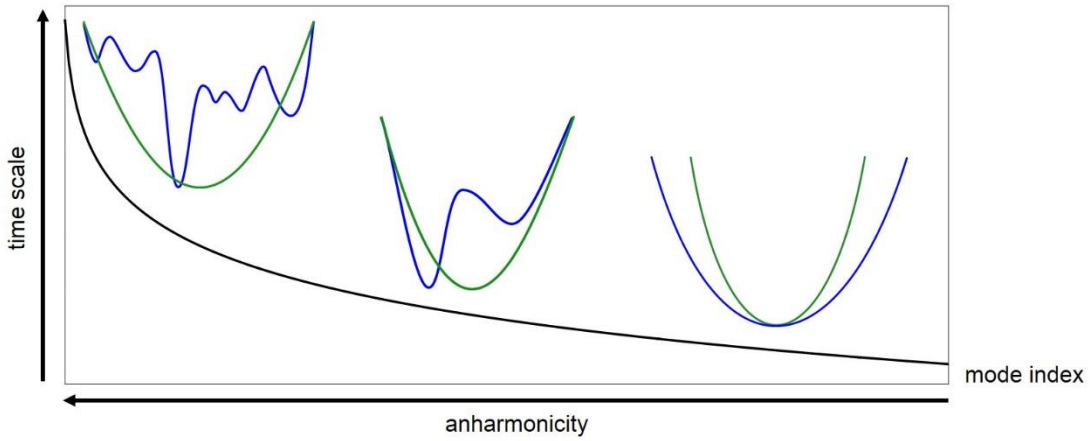

Fig 1. Time scales versus anharmonicity of the modes.

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 **Fig 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).

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In this study, we designed a novel approach using the Wiener–Khintchine theorem (WKT) [18] to define the "time period" for every principal component (PC) mode derived from PC analysis (PCA) of a long MD trajectory (**Fig 2**) [15,19,20]. To better describe a dynamics variable (e.g., order parameters), we identify a combination of PC modes whose fluctuation profile matches the variable the best (**Fig 3**) such that the time scale (see **S1 Fig**) of this dynamics variable can be indicated by the lowest PC mode among this set of modes. By analyzing several simulation trajectories of different lengths (120 to 600 ns) while gradually removing the slowest end of the spectrum, we obtain an estimated profile of order parameters that best agrees with the experimental ones [21]. The slowest end of the remaining PC modes infers a time period of 0.7 and 2.3 ns which clearly agrees with previously reported time ranges. Similarly, we determined the time scales for anisotropic displacement parameters (ADPs) [22] and the spatial spread of NMR-solved conformers [23] (**Fig 3**). From the theoretical end, through mapping the ENM modes to corresponding (an)harmonic PC modes and their WKT-inferred "time periods", we assign a time and size (σ^2) to each ENM mode (**Fig 4**) and obtain two power laws for the time scales and size of the modes, respectively, with respect to their eigenvalues. The newly established power laws thus turn ENM into a molecular timer and sizer, which, as showcased below, could temporally and spatially resolve the ratcheting, L1 stalk closing and head swiveling motions of the ribosome.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

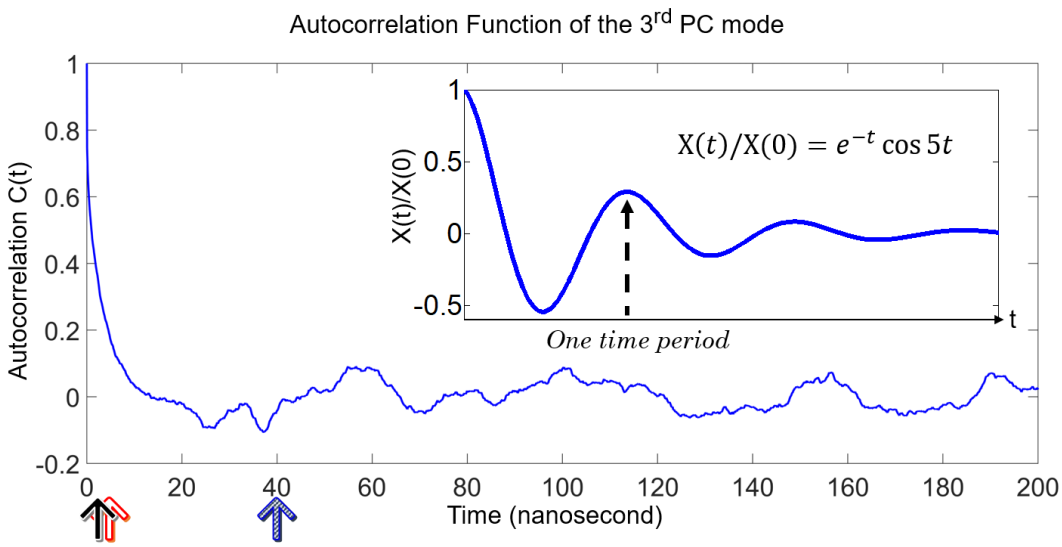

Fig 2. Defining the time scale of the first PC mode using the three methods.

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 [24], 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.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Determining the time scales of the slowest 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 [27] (PDB ID: 1UBQ). PCA analysis [20] 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 $\mathbf{Q}$ and decomposed into its corresponding eigenvalues and eigenvectors such that:

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

$\Delta\mathbf{R}$ 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 were projected onto all the PCs [15,20] to constitute a projection matrix $\mathbf{U}_{3N\times M}$ as

$$\mathbf{U} = \mathbf{V}^{\boldsymbol{T}}\mathbf{Q} \quad (2)$$

where the row k in $\mathbf{U}$, $\mathbf{u}_k = [u_{k,0}, u_{k,1}, \ldots, u_{k,M-1}]$, 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.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 [18], 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 **S1 Appendix**.

Fig 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 **Fig 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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transform of an exponential function is a Lorentzian function [24], 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 **Fig 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 **S1 Fig**. 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 [20], NMR order parameters [21] and x-ray crystallographical anisotropic displacement parameters (ADPs) [22]. Plotting the dynamical value for each residue against its residue index results in so-called "observed fluctuation profiles" (oFPs) (**Fig 3, top,** black circles). Every residue has one

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

magnitude of RMSF, one order parameter and 6 ADP values (including variance xx-, yy-, zz- and covariance xy-, yz- and xz- components [22]). 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)

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

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 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) [28]. 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 **S1 Appendix**). 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 **Fig 3A, top**. There are 210 ($3N - 6 = 210$) $\text{RMSF}_{MD,k}$ profiles, each calculated using all the PC modes $\geq k^{th}$ 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 (**Fig 3A, middle and bottom**).

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

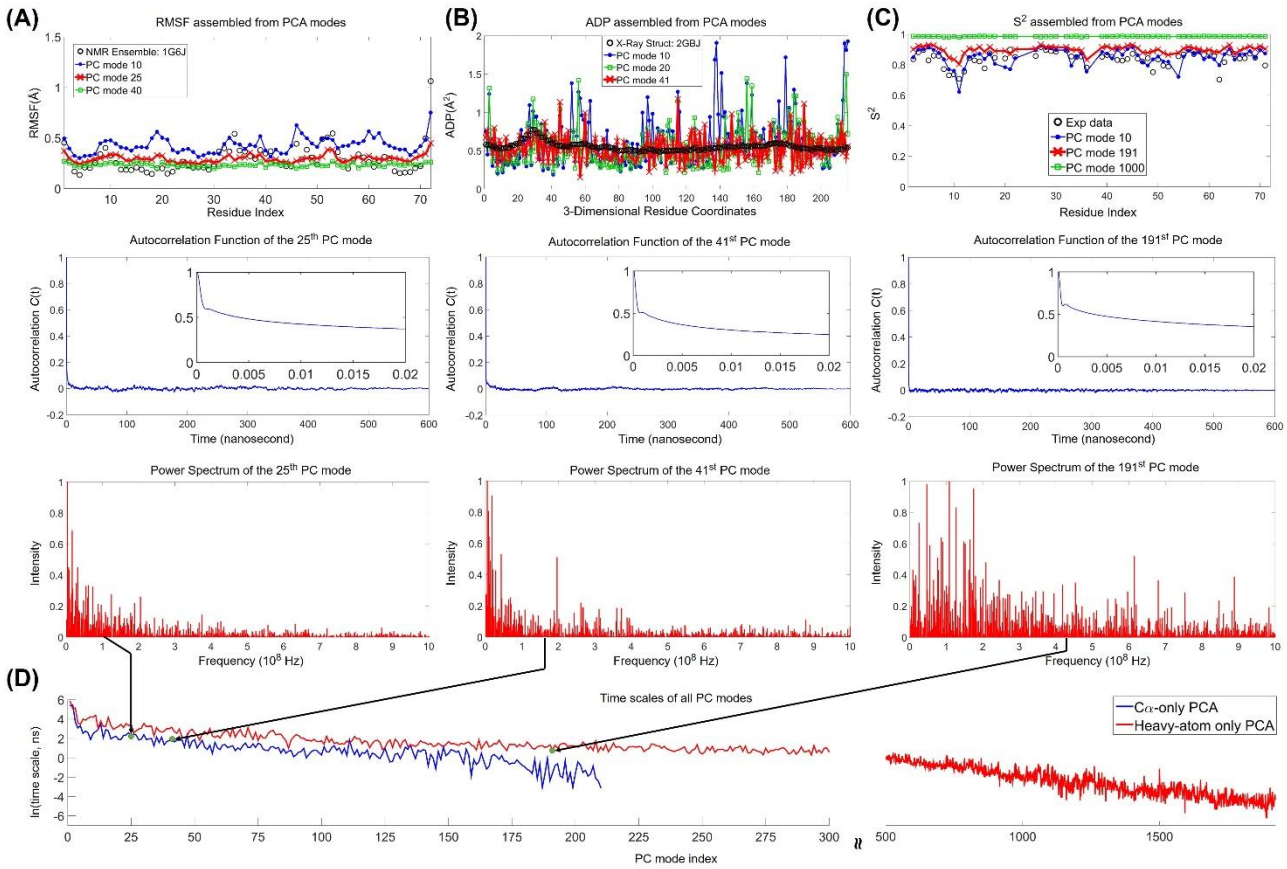

Fig 3. Time scale mapping for different experimental data that characterize protein dynamics.

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 (σ_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 **Eq 3**. In addition, **(A, bottom)** its corresponding

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 [21] 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 **Fig 3B, 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 **S2A Fig**.

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 [20,29,30]

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

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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (**Fig 3C** and see correlation as a function of PC mode index, peaked at the mode 191, in **S2B Fig**), characterizes the time scale and size of motion for the observed order parameter profiles [21]. 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 **S3A Fig**).

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) [31] 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 [32].

We observed that $S^2_{MD,k=1}$ shows slightly larger (less ordered) bond fluctuations than S^2_{exp} (**S3B Fig)**, 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.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 $\geq l$ (lower ANM modes have lower frequencies or smaller eigenvalues) was calculated as follows

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

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 **S1 Appendix** and

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig 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).

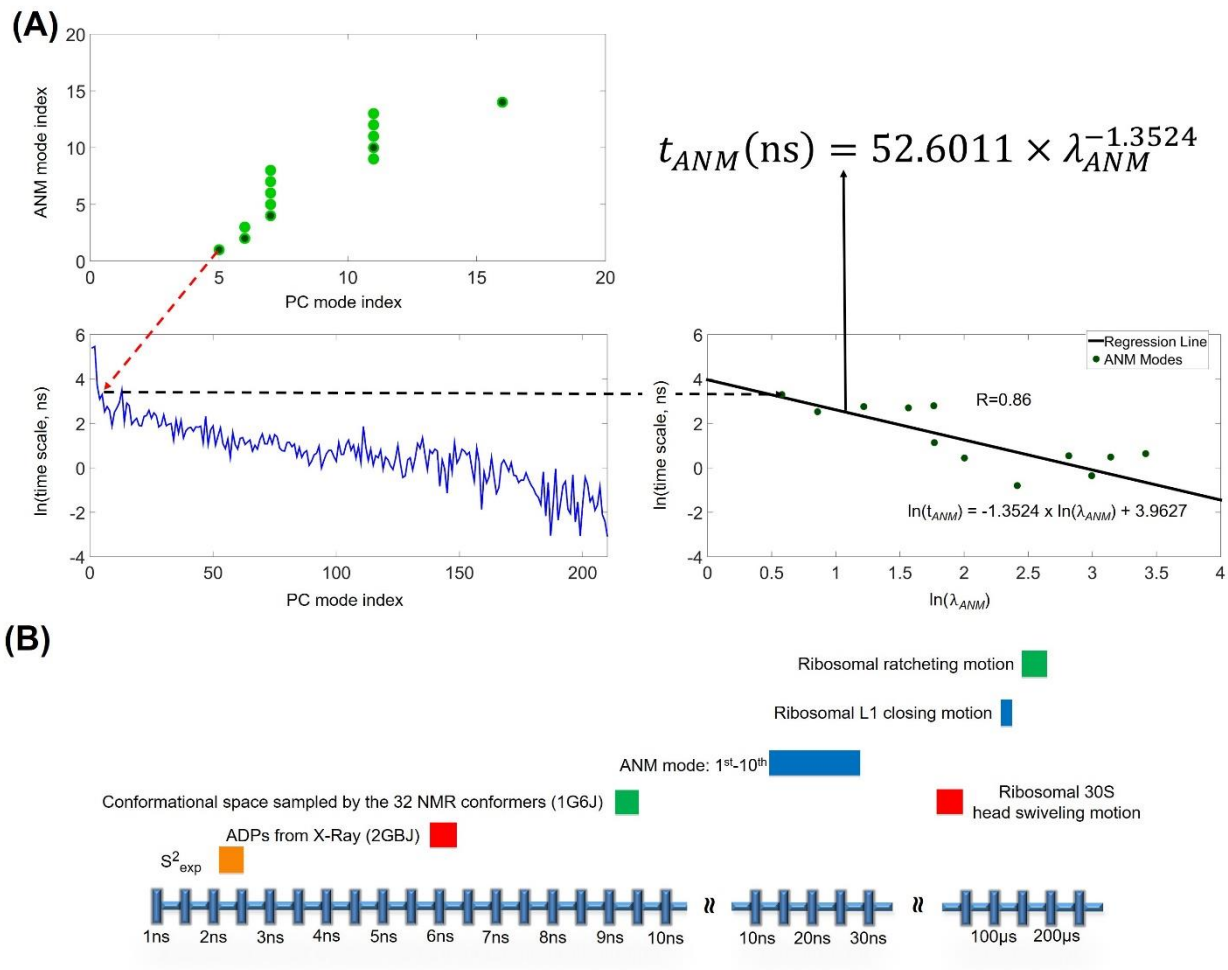

Fig 4. Frequency mapping for ANM to obtain the power law of τ_w as a function of ANM eigenvalues.

(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-,

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 **Fig 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 herein introduced PCA+WKT+FPM approach for ubiquitin, except that the time scales of ribosomal motions are estimated by the general power law $t_{ANM}(\mathrm{ns}) = 86.9387 \times \lambda_{ANM}^{-1.8886}$ derived from the combined τ_w and λ_{ANM} of ubiquitin, FGF2 and HPNAP.

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 (**Fig 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 **Fig 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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 $t_{ANM}(\mathrm{ns}) = 86.9387 \times \lambda_{ANM}^{-1.8886}$ (**S4A Fig**).

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 (**S4B Fig**), 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. **Fig 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 (**Figs 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

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 **S1 Appendix**). 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 $t \sim \langle\sigma^2\rangle$.

Predicting the size, conformation and time scale of ribosomal motions with the variance and time power laws

Finding the relevant mode that describes the ribosomal body rotation (ratcheting) motion. As an example to showcase a successful application of the power laws, we predict the size and

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

time scale of the ribosomal body rotation motion (ratcheting) [11,33] between the 30S and 50S subunits of the *Thermus thermophilus* ribosome [33], which is essential for ribosomal translocation during protein translation [34,35]. It is known that such a body rotation is generally around an axis that connects the mass centers of large and small subunits [11,33,36]. Therefore, we first performed ANM on the non-rotated structure of the ribosome (PDB ID: 4V6F [37]) and then created deformed conformations along each of the slowest 50 ANM modes (there are 63,774 modes in total) by assuming a unity value for the scaling factor $k_B T/\gamma$ (see **Eq S23 in S1 Appendix**). Next, we superimpose non-rotated and ANM-deformed ribosome only at their 50S subunits so that the combined rotation and deformation of 30S alone can be analyzed. We further isolated only the 30S rotation and obtained its corresponding rotation axis for each of the 50 ANM conformers (see **S1 Appendix**). Among these 50 modes, the *rotation axis* derived from mode 25 was identified to have the smallest deviation angle (θ = 11.6º) from the vector that connects the mass centers of 30S and 50S subunits. Therefore, we consider mode 25 as the ribosomal ratcheting motion.

Obtaining the size of functional motion and the rotated conformation of the ribosome. The variance of ANM mode 25 can be obtained from the variance power law $\sigma^2(\text{Å}^2) = 46.0538 \times \lambda_{ANM}^{-2.5085}$, the square root of which gives the size of conformational change, namely the structure difference between rotated and non-rotated ribosome. Given the unit vector of mode 25 and the size of the deformation, the rotated ribosome conformation can be obtained. Our predicted rotated ribosome conformation shows a 9° rotation which is close to the experimentally measured rotation of 7° (according to non-rotated state [PDB ID: 4V6F] and rotated state [PDB ID: 4V9H]) [37,38].

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Obtaining the time scale of ribosomal ratcheting motion. Using the introduced time power law and the ANM eigenvalues, one can estimate the time scales of these modes (**S1 Table**). The ribosomal ratcheting motion which mode 25 resembles is estimated to take place with a time scale of 160.7 μs, in contrast to 6.5 ns if estimated from a purely harmonic approximation [20] (by using PCA only).

It has been reported via the single-molecule Förster resonance energy transfer experiments [35] and stopped-flow apparatus [39] that the transit rate between the non-rotated and rotated state is faster than 25-100 ms [35] and 5 ms [39], respectively. Recently, it became possible to conduct explicit solvent simulations for the 70S ribosome [6,40]. A 1.3-μs explicit solvent MD simulation for the 70S ribosome [40] showed that 30S can have a body rotation of ~-2 degrees with the head swiveling approximately +5 degrees, which contrasts with the observed ~+7 degrees for a full body rotation and ~+20 degrees for a complete head rotation, characterized using X-ray [41] crystallography and cryo-EM [5,42]. In view of the portion of body rotation completed within 1.3 μs, the ANM power law estimates a time scale of ~161 μs, which falls in between incomplete body rotation (1.3 μs) and stopped flow suggested transition time (<5 ms).

Finding the relevant modes that describe 30S head swiveling and L1 stalk closing motion of the ribosome and characterizing the size and time scales of these motions. 30S head swiveling motion and L1 stalk motion are also known to be involved in the ribosomal translocation process [34]. The 30S head is defined as the 3' major domain of the 16S rRNA spanning residues 921-1396 and the complexed proteins [43] (colored in cyan at the bottom right of **S5 Fig**). Its rotational

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

motion relative to the body domain, coined as the swiveling motion, can swivel up to 18° after binding EF-G and the concomitant GTP hydrolysis. This intra-subunit motion together with the inter-subunit ratcheting of the whole 30S drives tRNA translocation [5]. Consequently, when the swiveling motion is impeded by an obstacle, such as an mRNA pseudoknot, the dissociation of E-site tRNA and EF-G can be hindered [44].

On the other hand, L1 stalk is defined as the complex between the helices 76-78 of 23S rRNA and the L1 ribosomal protein [45]. The L1 stalk motion is not only influenced by the 30S motions [45] and the downstream mRNA structures [46] but also by its binding to the P-site tRNA (L1 stalk closing) [47] and guiding it from the P-site to the E-site which involves motions as large as ~20 Å [48].

By applying the power laws, we are not only interested in reproducing observed sizes of conformational changes but also provided estimated timescales for these intrinsic and functional motions in ribosome.

A similar method used to identify the ANM mode that corresponds to the ratcheting motion was applied to find the corresponding mode for the swiveling motion (**S1 Appendix**). The main differences are that the axis of rotation for head swiveling is defined as the vector pointing from the COM of the 30S body to the COM of the 30S head and the deformed conformer is superimposed at the 50S and 30S body of the non-rotated ribosome (PDB ID: 4V6F), see **S1 Appendix** for details. As mentioned in the previous section, we can calculate the rotation axis for each of the 50 slowest ANM modes. The axis of rotation computed for ANM mode 32 has the

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

smallest angle (θ = 3.9°) deviating from the axis of rotation of the swiveling motion. Deforming the conformation along ANM mode 32 scaled by the variance power law results in an angle of rotation of 3.2° in contrast to the observed 4.4° rotation in the corresponding structures (PDB ID: 4V6F and PDB ID: 4V9H). The predicted time scale using the eigenvalue of mode 32 and the time power law is ~70 μs.

The L1 stalk interacts with the tRNA at the E-site of the ribosome during the ratcheting and the swiveling motions. To identify the ANM mode(s) corresponding to this motion, each of the top 50 slowest ANM modes was scaled by the variance power law and used to deform the non-rotated conformation (PDB ID: 4V6F), see **S1 Appendix** for details. 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 with the predicted time scales of ~105-128 μs. The obtained distance of ~26-30 Å is closer to the rotated conformation (25 Å; PDB ID: 4V9H) while the corresponding distance in the non-rotated ribosome (PDB ID: 4V6F) is 42 Å.

The fact that the timescale of the ratcheting motion is much closer to the L1 stalk closing than the swiveling motion suggests a stronger coupling between the ratcheting and the L1 stalk closing, whereas the 30S head moves more independently from the other two. Indeed, the correlation analysis over a dozen of cryo-EM and x-ray solved structures carried out by Agirrezabala et al. [49] showed that the correlation between the ratcheting angles and the swiveling angles is about 0.19 (*p-value* = 0.4241) while that between ratcheting and L1 stalk closing is 0.99 (p-value < 0.0001)

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Collectivity of the ribosomal motions. Tama and Sanejouand's definition of collectivity (how global a motion is) was applied to each of the ANM modes [50]. As expected, global ratcheting and L1 stalk closing (which is coupled to ratcheting) motions observed in ANM modes 25, 27 and 28 are among the modes with the highest collectivity within the slowest 50 ANM modes (**S5 Fig**), whereas the independent local motion of head swiveling has a relatively low collectivity.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

In this study, we show that the distributions (or profiles) of dynamical variables can serve as a hallmark to suggest the undelying time scale needed for such distributions. Using profile matching, the profile can be mapped to a specific ANM mode whose timescale can be estimated by a power law established in this study from long MD simulations, PCA, fluctuation profile-matching and the Wiener–Khintchine theorem.

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" 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 ($t \sim \langle\sigma^2\rangle$) [18]. 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 timer has now been implemented in the online server DynOmics [14] (http://dyn.life.nthu.edu.tw/oENM/) while the codes used for the analysis are available on https://github.com/Hong-Rui/bioStructureM.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

MD simulations

Detailed simulation protocols have been previously reported [25]. Briefly, MD simulations were performed using the PMEMD module of AMBER 10. The ff99SB force field [26] was used for the protein. 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. The system was 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. The long-range Coulomb energy was evaluated using the particle mesh Ewald (PME) method. MD simulation was conducted for 600 ns, and protein configurations were recorded every 0.1 ps for the analysis.

Acknowledgements

Part of the computational resources for this project are provided by National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) of Taiwan as well as supercomputers at the RCCS, The National Institute of Natural Science, and ISSP, The University of Tokyo.

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically Driven Protein Allostery. Nat Struct Mol Biol. 2006;13: 831–838. doi:10.1038/nsmb1132
2. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J Chem Theory Comput. 2012;8: 1542–1555. doi:10.1021/ct200909j
3. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. Science (80-). 2010;330: 341–346. doi:10.1126/science.1187409
4. Sanbonmatsu KY, Joseph S, Tung C-S. Simulating Movement of tRNA Into the Ribosome During Decoding. Proc Natl Acad Sci. 2005;102: 15854–15859. doi:10.1073/pnas.0503456102
5. Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature. 2010;468: 713–716. doi:10.1038/nature09547
6. Whitford PC, Onuchic JN, Sanbonmatsu KY. Connecting Energy Landscapes with Experimental Rates for Aminoacyl-tRNA Accommodation in the Ribosome. J Am Chem Soc. 2010;132: 13170–13171. doi:10.1021/ja1061399
7. Flory PJ, Gordon M, McCrum NG. Statistical Thermodynamics of Random Networks [and Discussion]. Proc R Soc A Math Phys Eng Sci. 1976;351: 351–380. doi:10.1098/rspa.1976.0146
8. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys Rev Lett. 1996;77: 1905–1908. doi:10.1103/PhysRevLett.77.1905

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9. Bahar I, Atilgan AR, Demirel MC, Erman B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. Phys Rev Lett. 1998;80: 2733–2736. doi:10.1103/PhysRevLett.80.2733

10. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. Biophys J. Elsevier; 2001;80: 505–515. doi:10.1016/S0006-3495(01)76033-X

11. Wang Y, Rader AJ, Bahar I, Jernigan RL. Global Ribosome Motions Revealed with Elastic Network Model. J Struct Biol. 2004;147: 302–314. doi:10.1016/j.jsb.2004.01.005

12. Li H, Chang Y-YY, Yang L-WW, Bahar I. iGNM 2.0: The Gaussian Network Model Database for Biomolecular Structural Dynamics. Nucleic Acids Res. 2016;44: D415-22. doi:10.1093/nar/gkv1236

13. Bahar I, Lezon TR, Yang L-W, Eyal E. Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys. 2010;39: 23–42. doi:10.1146/annurev.biophys.093008.131258

14. Li H, Chang Y-Y, Lee JY, Bahar I, Yang L-W. DynOmics: dynamics of structural proteome and beyond. Nucleic Acids Res. 2017;45: W374–W380. doi:10.1093/nar/gkx385

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: 447–472. doi:10.1016/0301-0104(91)87082-7

16. Kitao A, Go N. Investigating Protein Dynamics in Collective Coordinate Space. Curr Opin Struct Biol. 1999;9: 164–169. doi:10.1016/S0959-440X(99)80023-2

17. Yang L-W, Chng C-P. Coarse-Grained Models Reveal Functional Dynamics--I. Elastic

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Network Models--Theories, Comparisons and Perspectives. Bioinform Biol Insights. 2008;2: 25–45.

18. McQuarrie DA. The time-correlation function formalism II. Statistical Mechanics. University Science Books; 2000. pp. 543–592.

19. Amadei A, Linssen ABM, Berendsen HJC. Essential Dynamics of Proteins. Proteins Struct Funct Genet. 1993;17: 412–425. doi:10.1002/prot.340170408

20. 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: 606–14. doi:10.1093/bioinformatics/btp023

21. Tjandra N, Feller SE, Pastor RW, Bax A. Rotational Diffusion Anisotropy of Human Ubiquitin from 15N NMR Relaxation. J Am Chem Soc. 1995;117: 12562–12566. doi:10.1021/ja00155a020

22. 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: i175–i184. doi:10.1093/bioinformatics/btm186

23. Babu CR, Flynn PF, Wand AJ. Validation of Protein Structure from Preparations of Encapsulated Proteins Dissolved in Low Viscosity Fluids. J Am Chem Soc. 2001;123: 2691–2692. doi:10.1021/ja005766d

24. Boas ML. Mathematical Methods in the Physical Sciences. The Journal of Symbolic Logic. 2006. pp. 340–389.

25. Takemura K, Kitao A. Water Model Tuning for Improved Reproduction of Rotational Diffusion and NMR Spectral Density. J Phys Chem B. 2012;116: 6279–87. doi:10.1021/jp301100g

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26. 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: 712–725. doi:10.1002/prot.21123

27. Vijay-kumar S, Bugg CE, Cook WJ. Structure of Ubiquitin Refined at 1.8 Å Resolution. J Mol Biol. 1987;194: 531–544. doi:10.1016/0022-2836(87)90679-6

28. Pearson K. Note on Regression and Inheritance in the Case of Two Parents. Proc R Soc London. 1895;58: 240–242. doi:10.1098/rspl.1895.0041

29. Henry ER, Szabo A. Influence of Vibrational Motion on Solid State Line Shapes and NMR Relaxation. J Chem Phys. 1985;82: 4753. doi:10.1063/1.448692

30. Best RB, Vendruscolo M. Determination of Protein Structures Consistent with NMR Order Parameters. J Am Chem Soc. 2004;126: 8090–8091. doi:10.1021/ja0396955

31. Lee AL, Flynn PF, Wand AJ. 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: 2891–2902. doi:10.1021/ja983758f

32. Nederveen AJ, Bonvin AMJJ. NMR Relaxation and Internal Dynamics of Ubiquitin from a 0.2 μs MD Simulation. J Chem Theory Comput. 2005;1: 363–74. doi:10.1021/ct0498829

33. Tama F, Valle M, Frank J, Brooks CL. Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy. Proc Natl Acad Sci. 2003;100: 9319–9323. doi:10.1073/pnas.1632476100

34. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature. 2000;406: 318–22. doi:10.1038/35018597

35. Cornish P V., Ermolenko DN, Noller HF, Ha T. Spontaneous Intersubunit Rotation in Single Ribosomes. Mol Cell. 2008;30: 578–588. doi:10.1016/j.molcel.2008.05.004

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

36. Kurkcuoglu O, Doruker P, Sen TZ, Kloczkowski A, Jernigan RL. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. Phys Biol. 2008;5: 046005. doi:10.1088/1478-3975/5/4/046005

37. Jenner LB, Demeshkina N, Yusupova G, Yusupov M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. Nat Struct Mol Biol. 2010;17: 555–60. doi:10.1038/nsmb.1790

38. Tourigny DS, Fernandez IS, Kelley AC, Ramakrishnan V. Elongation Factor G Bound to the Ribosome in an Intermediate State of Translocation. Science (80-). 2013;340: 1235490–1235490. doi:10.1126/science.1235490

39. Guo Z, Noller HF. Rotation of the Head of the 30S Ribosomal Subunit During mRNA Translocation. Proc Natl Acad Sci. 2012;109: 20391–20394. doi:10.1073/pnas.1218999109

40. Whitford PC, Blanchard SC, Cate JHD, Sanbonmatsu KY. Connecting the Kinetics and Energy Landscape of tRNA Translocation on the Ribosome. Chen S-J, editor. PLoS Comput Biol. 2013;9: e1003003. doi:10.1371/journal.pcbi.1003003

41. Zhang W, Dunkle JA, Cate JHD. Structures of the Ribosome in Intermediate States of Ratcheting. Science. 2009;325: 1014–7. doi:10.1126/science.1175275

42. Fischer N, Konevega AL, Wintermeyer W, Rodnina M V, Stark H. Ribosome Dynamics and tRNA Movement by Time-Resolved Electron Cryomicroscopy. Nature. Nature Publishing Group; 2010;466: 329–333. doi:10.1038/nature09206

43. Mohan S, Donohue JP, Noller HF. Molecular mechanics of 30S subunit head rotation. Proc Natl Acad Sci. 2014;111: 13325–13330. doi:10.1073/pnas.1413731111

44. Caliskan N, Katunin VI, Belardinelli R, Peske F, Rodnina M V. Programmed -1 frameshifting by kinetic partitioning during impeded translocation. Cell. Elsevier Inc.;

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2014;157: 1619–1631. doi:10.1016/j.cell.2014.04.041

45. Fei J, Kosuri P, MacDougall DD, Gonzalez RL. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. Mol Cell. 2008;30: 348–59. doi:10.1016/j.molcel.2008.03.012

46. Chen C, Zhang H, Broitman SL, Reiche M, Farrell I, Cooperman BS, et al. Dynamics of translation by single ribosomes through mRNA secondary structures. Nat Struct Mol Biol. 2013;20: 582–8. doi:10.1038/nsmb.2544

47. Trabuco LG, Schreiner E, Eargle J, Cornish P, Ha T, Luthey-Schulten Z, et al. The role of L1 stalk-tRNA interaction in the ribosome elongation cycle. J Mol Biol. 2010;402: 741–60. doi:10.1016/j.jmb.2010.07.056

48. Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J. Locking and unlocking of ribosomal motions. Cell. 2003;114: 123–34. Available: http://www.ncbi.nlm.nih.gov/pubmed/12859903

49. Agirrezabala X, Liao HY, Schreiner E, Fu J, Ortiz-Meoz RF, Schulten K, et al. Structural characterization of mRNA-tRNA translocation intermediates. Proc Natl Acad Sci U S A. 2012;109: 6094–9. doi:10.1073/pnas.1201288109

50. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng. 2001;14: 1–6. doi:10.1093/protein/14.1.1

bioRxiv preprint doi: https://doi.org/10.1101/384511; this version posted November 2, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Information

S1 Fig. Time scale of each PC mode of ubiquitin estimated using the three tested methods.

S2 Fig. The PC modes that best match the dynamics variables of ubiquitin.

S3 Fig. NMR order parameters (S^2) estimated from the 120 ns and 600 ns MD trajectories.

Fig S4. Comparing IWPs to power law estimated periods and plotting eigenvalues of ANM's slowest modes to the lengths of proteins.

Fig S5. The collectivity (κ) of the top 50 slowest ANM modes where four of which are the discussed functional motions in ribosome.

S1 Table. The top 50 slowest ribosomal motions characterized by ANM.

S1 Appendix. Supporting methods and supporting references.