Models with Energy Penalty on Interresidue Rotation Address Insufficiencies of Conventional Elastic Network Models

Lee-Wei Yang

Supporting Material

Table S1. Correlation of B-factors and predictions by simple EN models for a set of 30 protein structures of ultra-high resolutions

PDB ID	Length	$Corr^{I}_{GNM}$	$Corr^{I}_{EPIRM}$	$Corr^{I}_{ANM}$	$Corr^{A}_{EPIRM}$	$Corr^{A}_{ANM}$
1a6m	151	0.653	0.656	0.664	0.488	0.552
1f9y	158	0.716	0.714	0.645	0.545	0.426
1g66	207	0.541	0.543	0.548	0.429	0.48
1g6x	58	0.789	0.783	0.758	0.644	0.686
1gdn	227	0.593	0.595	0.585	0.438	0.464
1i1w	303	0.641	0.650	0.651	0.489	0.559
1i1x	302	0.806	0.811	0.803	0.607	0.667
1k6u	58	0.826	0.823	0.780	0.628	0.664
1kms	184	0.436	0.435	0.416	0.359	0.349
1kmv	185	0.437	0.437	0.419	0.345	0.384
1kt5	175	0.639	0.697	0.647	0.614	0.558
1lok	291	0.609	0.606	0.573	0.528	0.525
1lug	259	0.732	0.729	0.660	0.56	0.554
1m1q	90	0.685	0.689	0.649	0.423	0.401
1m40	263	0.636	0.635	0.590	0.477	0.475
1me3	215	0.651	0.649	0.587	0.481	0.472
1me4	215	0.657	0.654	0.594	0.514	0.491
1nwz	125	0.431	0.430	0.475	0.364	0.425
1nym	263	0.661	0.663	0.613	0.445	0.506
1oc7	364	0.786	0.784	0.751	0.747	0.689
1oq5	256	0.728	0.724	0.681	0.538	0.542
1pwm	316	0.658	0.661	0.563	0.527	0.464
1q0n	158	0.587	0.583	0.551	0.447	0.444
1q2q	359	0.567	0.570	0.584	0.476	0.563
1rtq	291	0.701	0.700	0.698	0.556	0.591
1swz	164	0.564	0.561	0.592	0.445	0.564
1sx7	164	0.578	0.579	0.621	0.456	0.585
1z8a	315	0.624	0.628	0.570	0.524	0.495
3lzt	129	0.733	0.734	0.604	0.545	0.443
4lzt	129	0.639	0.643	0.469	0.489	0.437
mean		**0.643±0.019**	**0.643±0.019**	**0.610±0.017**	**0.504±0.016**	**0.515±0.016**

The 30 non-homologous high-resolution (< 1Å) proteins are taken from a set reported previously (14). Due to the high resolution, anisotropic displacement parameters (ADPs) are available to all the listed proteins along with their corresponding temperature factors (B-factors). '$Corr^{I}$' is the correlation between ENM predictions and B-factors, and '$Corr^{A}$' is the correlation between these predictions and ADPs. Note that GNM is an isotropic model and

therefore provides no '$Corr^A$' results. Paired student's t-tests reveal superior performance of GNM and EPIRM in predicting B-factors to that of ANM (see the average $Corr^I$ values). Yet, the data suggest that EPRIM and ANM describe ADPs equally well (per the statistically identical $Corr^A$ averages). Means and standard errors are listed in the bottom of the table. Here, 15Å cutoff distance is taken for all the 3 ENMs to maintain an equal sparsity in Hessians.

Table S2. Correlation of C_α distributions and predictions by EPIRM and ANM for a set of 64 NMR-determined structural ensembles

PDB ID	Length	$Corr^I_{EPIRM}$	$Corr^I_{ANM}$	$Corr^A_{EPIRM}$	$Corr^A_{ANM}$
1a63	130	0.865	0.869	0.763	0.789
1a67	108	0.554	0.505	0.497	0.478
1ah2	269	0.769	0.761	0.675	0.702
1ao8	163	0.683	0.730	0.596	0.673
1ax3	162	0.899	0.737	0.872	0.693
1ayk	169	0.815	0.769	0.750	0.686
1b6f	159	0.846	0.947	0.791	0.909
1b8q	134	0.832	0.659	0.717	0.577
1bip	122	0.898	0.841	0.838	0.760
1bsh	138	0.760	0.866	0.496	0.846
1bvm	123	0.532	0.434	0.444	0.391
1byn	128	0.616	0.521	0.562	0.488
1clh	166	0.722	0.748	0.666	0.705
1cur	155	0.576	0.598	0.517	0.541
1cye	129	0.740	0.477	0.679	0.448
1df3	162	0.561	0.409	0.522	0.383
1e1g	104	0.849	0.792	0.656	0.737
1e8l	129	0.502	0.534	0.397	0.480
1eky	129	0.686	0.680	0.581	0.633
1eq0	158	0.686	0.868	0.588	0.819
1ezo	370	0.747	0.696	0.451	0.579
1fhb	108	0.805	0.537	0.640	0.465
1fhs	112	0.820	0.778	0.681	0.685
1fr0	125	0.868	0.954	0.810	0.827
1g9e	117	0.713	0.635	0.607	0.632
1gd5	130	0.716	0.793	0.663	0.749
1gio	125	0.830	0.905	0.766	0.824
1go0	102	0.871	0.882	0.837	0.823
1i4v	230	0.940	0.841	0.911	0.801
1iku	188	0.738	0.785	0.629	0.735
1inz	148	0.931	0.916	0.902	0.742
1ivt	122	0.675	0.687	0.617	0.626
1j1h	123	0.776	0.842	0.701	0.734
1k19	112	0.733	0.458	0.678	0.427
1klv	100	0.858	0.932	0.821	0.891
1ktm	139	0.902	0.749	0.783	0.657
1l3g	123	0.943	0.784	0.907	0.763

1ls4	164	0.719	0.885	0.645	0.815
1m0v	137	0.751	0.757	0.677	0.699
1mm4	170	0.859	0.923	0.808	0.822
1mph	106	0.619	0.737	0.570	0.659
1mvg	125	0.712	0.240	0.597	0.325
1nmv	163	0.851	0.789	0.781	0.667
1orm	148	0.775	0.822	0.528	0.707
1qce	369	0.426	0.769	0.263	0.761
1qn0	112	0.529	0.761	0.485	0.635
1qnd	123	0.825	0.802	0.772	0.727
1qtt	117	0.837	0.948	0.764	0.757
1r3b	202	0.733	0.886	0.602	0.651
1rch	155	0.670	0.755	0.575	0.659
1rck	106	0.710	0.679	0.640	0.624
1ry4	128	0.933	0.926	0.814	0.851
1sjg	112	0.424	0.438	0.365	0.388
1soy	106	0.532	0.690	0.420	0.624
1svq	94	0.647	0.498	0.579	0.474
1t3v	124	0.938	0.614	0.886	0.445
1tr4	226	0.507	0.635	0.398	0.559
1vre	147	0.519	0.605	0.403	0.541
1xo4	103	0.880	0.821	0.813	0.712
1xpw	143	0.750	0.850	0.699	0.769
1y8b	723	0.478	0.542	0.406	0.476
2prf	125	0.647	0.603	0.533	0.533
3msp	252	0.896	0.908	0.473	0.740
3phy	125	0.751	0.821	0.655	0.725
mean		0.737±0.017	0.728±0.020	0.643±0.019	0.657±0.018

The size of C_α distribution in a given NMR ensemble is manifested by its inter-conformer r.m.s.d. such that $\mathbf{rmsd}_i = \sqrt{\sum_{k=1}^{m}(\mathbf{r}_{i,k} - \bar{\mathbf{r}}_i)/m}$ where $\bar{\mathbf{r}}_i = \sum_{k=1}^{m}\mathbf{r}_{i,k}/m$ that is the average position of the *i*-th C_α over *m* models, and $\mathbf{r}_{i,k}$ is the position vector of the *i*-th C_α residue in the *k*th conformer. Note that $\mathbf{rmsd}_i$ is a 3-d vector containing x-, y- and z-components. Before applying the above calculations, all the conformers are iteratively best-aligned against their mean position until it converges (8). $\mathbf{Corr}^{\mathbf{I}}$ is the correlation between $|\mathbf{rmsd}_i|$ (i=1..*N*) and the fluctuation magnitude predicted for all the *N* C_αs by EPIRM and ANM (*N* is the length of the protein). On the other hand, $\mathbf{Corr}^{\mathbf{A}}$ is the correlation between $\mathbf{rmsd}_i$ (i=1..*N*) and the fluctuation magnitude predicted for each x-, y- and z-components of all the *N* C_αs. Means and standard errors are listed in the last row of the table. Paired student's t-tests reveal $\mathbf{Corr}^{\mathbf{I}}_{\mathbf{EPIRM}} = \mathbf{Corr}^{\mathbf{I}}_{\mathbf{ANM}} > \mathbf{Corr}^{\mathbf{A}}_{\mathbf{EPIRM}} = \mathbf{Corr}^{\mathbf{A}}_{\mathbf{ANM}}$ using a standard 5% threshold. The 64 non-homologous NMR ensembles are taken from our previous study (27) where we show that GNM gives a correlation of 0.746 ± 0.017 ($\mathbf{Corr}^{\mathbf{I}}_{\mathbf{GNM}}$) over the same set. Here, the ENM calculations are performed on the first conformer deposited in each ensemble. Although different conformer can as well be used, our former study demonstrated this does not alter the results statistically (27).

Table S3. **Correlations between dominant modes and open↔closed transitions in the three ENMs, when either 'open' or 'closed' structures are used as the input.**

Protein	Open/Closed	EPIRM	ANM	eGNM
		α^o/α^c (k_o*/ k_c*)	α^o/α^c (k_o*/ k_c*)	α^o/α^c (k_o*/ k_c*)
Calmodulin	1cll/1ctr	0.56/0.44 (1/1)	0.48/0.39 (6/3)	0.46/0.38 (1/1)
Diphtheria toxin	1ddt/1mdt	0.65/0.51 (1/4)	0.48/0.35 (1/5)	0.52/0.46 (1/4)
LAO binding	2lao/1lst	0.46/0.47 (1/15)	0.76/0.56 (1/4)	0.40/0.45 (6/15)
Enolase	3enl/7enl	0.38/0.30 (1/1)	0.31/0.25 (1/2)	0.28/0.27 (1/1)
Adenylate Kinase	4akeB/1e4vA	0.71/0.43 (1/1)	0.79/0.54 (1/1)	0.59/0.40 (1/1)
Thymidylate synthase	3tms/2tsc	0.48/0.46 (1/2)	0.44/0.22 (6/8)	0.43/0.46 (1/2)
DHFR	5dfr/4dfr	0.45/0.40 (1/1)	0.62/0.66 (1/1)	0.32/0.37 (1/12)
Citrate synthase	5csc/6csc	0.75/0.50 (1/1)	0.89/0.39 (3/5)	0.71/0.50 (1/1)
Yhdh	1o89A/1o8cB	0.62/0.53 (1/1)	0.46/0.41(3/1)	0.40/0.34 (1/1)
Actin-Related Protein	1k8k/1tyq	0.64/0.67 (1/1)	0.73/0.56 (2/6)	0.52/0.60 (1/1)
Average		**0.57**/0.47(**1.0**/2.8)	**0.60**/0.43(**2.5**/3.6)	**0.46**/0.42(**1.5**/3.9)

PDB structures used for open/closed conformation are listed in the second column. In three examined ENMs (column 3-5), k_o* is the internal mode that has the highest correlation (indicated by α^o) with observed open↔closed transitions when 'open (unbound)' structures are used as the ENM input, whereas k_c* is the mode that produces the highest correlation α^c with the observed conformational changes when 'closed (inhibitor-bound)' structures are used as the input. For instance, the conformational transition for Calmodulin from 'open' (1cll) to 'closed' (1ctr) state can be best described by the 6th ANM mode with a 0.48 correlation when the 'open' structure is used, or by the 3rd ANM mode with a 0.39 correlation when the 'closed' structure is used. Here the paired student T-test is used to examine the statistical equivalence or disparity between ENM groups. As 'open' conformation is used as the input, EPIRM and ANM have statistically identical averages (0.57 and 0.60 respectively) while outperforming eGNM that has an average of 0.46. When 'closed' conformation is used as the input, the three ENMs have statistically equivalent performances despite the higher mean value of EPIRM (0.47).

I. Detailed derivations of the eGNM Hessian, $\Gamma \otimes \mathbf{I}$ matrix, from the GNM potential

The GNM energy E_{GNM} takes the form

$$E_{GNM} = \sum_{i,j=1}^{N} \frac{\gamma}{2} (\vec{\mathbf{r}}_{ij} - \vec{\mathbf{r}}_{ij}^{\,0}) \bullet (\vec{\mathbf{r}}_{ij} - \vec{\mathbf{r}}_{ij}^{\,0}) H(R_c - |\vec{\mathbf{r}}_{ij}^{\,0}|) \qquad \text{(SI-1)}$$

Here $\vec{\mathbf{r}}_{ij}$ and $\vec{\mathbf{r}}_{ij}^{\,0}$ are the positional vectors pointing from the C_α atom i to the C_α atom j at an instantaneous moment and at the equilibrium state (readily obtained from solved X-ray or NMR structures) respectively. The Heaviside step function in the end of right-hand side ensures the values are non-vanishing only when the equilibrium departure of atom pairs is within a cutoff distance R_c. The difference of vectors $\vec{\mathbf{r}}_{ij}$ and $\vec{\mathbf{r}}_{ij}^{\,0}$ can be expressed as the difference between vectors $\Delta\vec{\mathbf{r}}_j$ and $\Delta\vec{\mathbf{r}}_i$, the deviation of atom i and j from their mean values, such that

$$E_{GNM} = \sum_{i,j=1}^{N} \frac{\gamma}{2} (\Delta\vec{\mathbf{r}}_j - \Delta\vec{\mathbf{r}}_i) \bullet (\Delta\vec{\mathbf{r}}_j - \Delta\vec{\mathbf{r}}_i) H(R_c - |\vec{\mathbf{r}}_{ij}^{\,0}|) \qquad \text{(SI-2)}$$

A schematic illustration for the relation of these vectors can be found in Figure S1.
The matrix-vector form of the same expression reads as

$$\begin{aligned} E_{GNM} &= \frac{\gamma}{2} \Delta\mathbf{R}^T (\mathbf{\Gamma} \otimes \mathbf{I}) \Delta\mathbf{R} \\ &= \frac{\gamma}{2} (\Delta\mathbf{X}^T \mathbf{\Gamma} \Delta\mathbf{X} + \Delta\mathbf{Y}^T \mathbf{\Gamma} \Delta\mathbf{Y} + \Delta\mathbf{Z}^T \mathbf{\Gamma} \Delta\mathbf{Z}) \end{aligned} \qquad \text{(SI-3)}$$

$\Delta\mathbf{R}$ is a $3N$-d column vector and $\Delta\mathbf{X}$, $\Delta\mathbf{Y}$ and $\Delta\mathbf{Z}$ are column vectors of N dimension.
where $\Delta\mathbf{R}^{\mathrm{T}} = (\Delta x_1, \Delta y_1, \Delta z_1, \Delta x_2, \Delta y_2, \Delta z_2, \ldots \Delta x_N, \Delta y_N, \Delta z_N)$

$\Delta\mathbf{X}^{\mathrm{T}} = (\Delta x_1, \Delta x_2 \ldots \Delta x_N)$; $\Delta\mathbf{Y}^{\mathrm{T}} = (\Delta y_1, \Delta y_2 \ldots \Delta y_N)$; $\Delta\mathbf{Z}^{\mathrm{T}} = (\Delta z_1, \Delta z_2 \ldots \Delta z_N)$;

$\Delta\alpha_i$ is the positional deviation from the mean for node i at a instant moment; α is the Cartesian coordinate of x-, y- or z- of the node i.

$\mathbf{\Gamma}$ is a $N \times N$ connectivity matrix, of which the off-diagonal elements $\Gamma_{ij} = -1$ if i is in contact with j within a cutoff distance and zero otherwise. The diagonal elements $\Gamma_{ii} = -\sum_{j=1, j\neq i}^{N} \Gamma_{ij}$; $\mathbf{\Gamma} \otimes \mathbf{I}$ is a $3N \times 3N$ matrix comprises $N \times N$ super-elements, each of which is a 3×3 matrix for pair ij and the super-element ij takes values $\Gamma_{ij}\ \mathbf{I}$ where $\mathbf{I}$ is a 3×3 identity matrix. $\mathbf{\Gamma} \otimes \mathbf{I}$ *is the Hessian for GNM potential*. The elements in $\mathbf{\Gamma} \otimes \mathbf{I}$ are the second derivatives of the E_{GNM}. The proof of which is shown in the Supplementary II.

The derivation of the section starts from the definition of *ensemble average* (or expectation value) of a certain physical quantity represented as a random variable; the *probability* associated with each instantaneous value of such quantity can be defined by the *potential* of the system at the

value through Boltzmann relation. The potential featured in GNM here is a simple, residue-based, pairwise potential.

For an instantaneous conformation, its probability assumes the Boltzmann relation such that

$$\begin{aligned}\Pr(\Delta\mathbf{R}) &= \exp(-E_{GNM}(\Delta\mathbf{R})/k_BT)\\ &= \exp(-\frac{\gamma}{2}(\Delta\mathbf{X}^T\boldsymbol{\Gamma}\Delta\mathbf{X}+\Delta\mathbf{Y}^T\boldsymbol{\Gamma}\Delta\mathbf{Y}+\Delta\mathbf{Z}^T\boldsymbol{\Gamma}\Delta\mathbf{Z})/k_BT)\\ &= \exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{X}^T\boldsymbol{\Gamma}\Delta\mathbf{X})\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{Y}^T\boldsymbol{\Gamma}\Delta\mathbf{Y})\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{Z}^T\boldsymbol{\Gamma}\Delta\mathbf{Z})\\ &= \Pr(\Delta\mathbf{X})\Pr(\Delta\mathbf{Y})\Pr(\Delta\mathbf{Z}) \qquad \text{(SI-4)}\end{aligned}$$

Let us consider a $N\times N$ residue-residue (or node-node) correlation matrix $\mathbf{C}$, of which the component C_{ij} is $\Delta\vec{\mathbf{r}}_i \bullet \Delta\vec{\mathbf{r}}_j$. It is clear that

$$<\mathbf{C}> \;=\; <\Delta\mathbf{X}^T\Delta\mathbf{X}> + <\Delta\mathbf{Y}^T\Delta\mathbf{Y}> + <\Delta\mathbf{Z}^T\Delta\mathbf{Z}> \qquad \text{(SI-5)}$$

$$<\mathbf{C}> = \frac{\int \mathbf{C}\Pr(\Delta\mathbf{R})\,\mathrm{d}\Delta\mathbf{R}}{\int \Pr(\Delta\mathbf{R})\,\mathrm{d}\Delta\mathbf{R}} \qquad \text{(SI-6)}$$

$$<\Delta\mathbf{X}^T\Delta\mathbf{X}> = \frac{\int \Delta\mathbf{X}^T\Delta\mathbf{X}\Pr(\Delta\mathbf{R})\,\mathrm{d}\Delta\mathbf{R}}{\int \Pr(\Delta\mathbf{R})\,\mathrm{d}\Delta\mathbf{R}} = \frac{\int \Delta\mathbf{X}^T\Delta\mathbf{X}\Pr(\Delta\mathbf{X})\Pr(\Delta\mathbf{Y})\Pr(\Delta\mathbf{Z})\,\mathrm{d}\Delta\mathbf{X}\,\mathrm{d}\Delta\mathbf{Y}\,\mathrm{d}\Delta\mathbf{Z}}{\int \Pr(\Delta\mathbf{X})\Pr(\Delta\mathbf{Y})\Pr(\Delta\mathbf{Z})\,\mathrm{d}\Delta\mathbf{X}\,\mathrm{d}\Delta\mathbf{Y}\,\mathrm{d}\Delta\mathbf{Z}} =$$

$$\frac{\int_{-\infty}^{\infty}\Delta\mathbf{X}^T\Delta\mathbf{X}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{X}^T\boldsymbol{\Gamma}\Delta\mathbf{X})d\Delta\mathbf{X}\int_{-\infty}^{\infty}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{Y}^T\boldsymbol{\Gamma}\Delta\mathbf{Y})d\Delta\mathbf{Y}\int_{-\infty}^{\infty}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{Z}^T\boldsymbol{\Gamma}\Delta\mathbf{Z})d\Delta\mathbf{Z}}{\int_{-\infty}^{\infty}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{X}^T\boldsymbol{\Gamma}\Delta\mathbf{X})d\Delta\mathbf{X}\int_{-\infty}^{\infty}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{Y}^T\boldsymbol{\Gamma}\Delta\mathbf{Y})d\Delta\mathbf{Y}\int_{-\infty}^{\infty}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{Z}^T\boldsymbol{\Gamma}\Delta\mathbf{Z})d\Delta\mathbf{Z}}$$

$$= \frac{\int_{-\infty}^{\infty}\Delta\mathbf{X}\,\Delta\mathbf{X}^T\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{X}^T\Gamma\Delta\mathbf{X})\,\mathrm{d}\Delta\mathbf{X}}{\int_{-\infty}^{\infty}\exp(-\frac{\gamma}{2k_BT}\Delta\mathbf{X}^T\Gamma\Delta\mathbf{X})\,\mathrm{d}\Delta\mathbf{X}} \qquad \text{(SI-7)}$$

The Gaussian integrals in the denominator and the integral in the numerator can be solved using the equalities $\int_{-\infty}^{\infty}e^{-ax^2}dx = \sqrt{(\frac{\pi}{a})}$ and $\int_{-\infty}^{\infty}x^2e^{-ax^2}dx = \frac{\sqrt{\pi}}{2}a^{-3/2}$, we therefore obtain

$<\Delta\mathbf{X}\Delta\mathbf{X}^T> = \frac{k_BT}{\gamma}\boldsymbol{\Gamma}^{-1}$; Similarly, $<\Delta\mathbf{Y}\Delta\mathbf{Y}^T> = \frac{k_BT}{\gamma}\boldsymbol{\Gamma}^{-1}$ and $<\Delta\mathbf{Z}\Delta\mathbf{Z}^T> = \frac{k_BT}{\gamma}\boldsymbol{\Gamma}^{-1}$, and

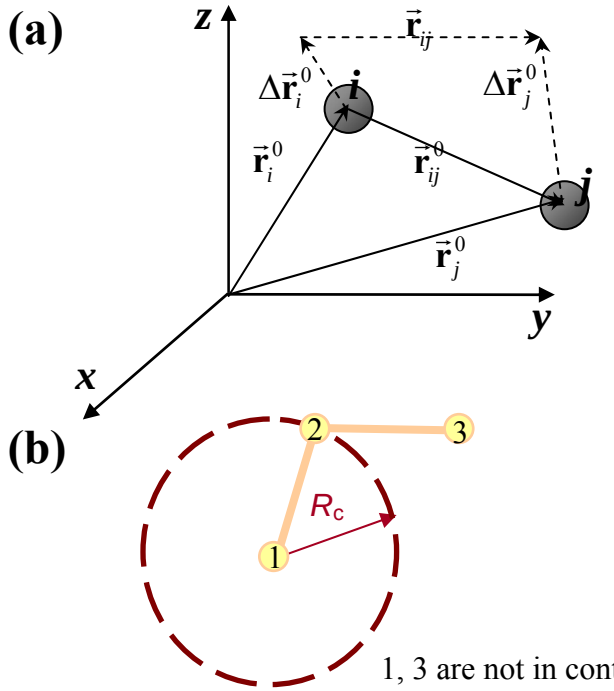
$$<\mathbf{C}> \;=\; <\Delta\mathbf{X}^T\Delta\mathbf{X}> + <\Delta\mathbf{Y}^T\Delta\mathbf{Y}> + <\Delta\mathbf{Z}^T\Delta\mathbf{Z}> = \frac{3k_BT}{\gamma}\boldsymbol{\Gamma}^{-1} \qquad \text{(SI-8)}$$

Note that here the imposed condition $<\Delta\mathbf{X}\Delta\mathbf{X}^{\boldsymbol{T}}> = <\Delta\mathbf{Y}\Delta\mathbf{Y}^{\boldsymbol{T}}> = <\Delta\mathbf{Z}\Delta\mathbf{Z}^{\boldsymbol{T}}>$ is not an assumption if GNM potential is so assumed in the Eq.SI-1. However, in Flory's treatment, polymer chain is assumed Gaussian in the first place (10).

Also, since $E_{GNM} = \frac{\gamma}{2}\Delta\mathbf{R}^T(\mathbf{\Gamma}\otimes\mathbf{I})\Delta\mathbf{R}$, it can be shown in a similar way that

$$\Pr(\Delta\mathbf{R}) = \exp(-\frac{\gamma}{2k_BT}\Delta R^T(\mathbf{\Gamma}\otimes\mathbf{I})\Delta R) \qquad \text{(SI-9)}$$

$$<\Delta\mathbf{R}^T\Delta\mathbf{R}> = \frac{\int \Delta\mathbf{R}^T\Delta\mathbf{R}\Pr(\Delta\mathbf{R})\,\mathrm{d}\,\Delta\mathbf{R}}{\int \Pr(\Delta\mathbf{R})\,\mathrm{d}\,\Delta\mathbf{R}} = \frac{k_BT}{\gamma}(\mathbf{\Gamma}\otimes\mathbf{I})^{-1} \qquad \text{(SI-10)}$$



1, 3 are not in contact for a given cutoff R_c

$$E_{GNM} = \gamma/2\ (\Delta\vec{\mathbf{r}}_{12}{}^2 + \Delta\vec{\mathbf{r}}_{23}{}^2)$$

$$= \gamma/2[(\Delta\vec{\mathbf{r}}_2 - \Delta\vec{\mathbf{r}}_1)^2 + (\Delta\vec{\mathbf{r}}_3 - \Delta\vec{\mathbf{r}}_2)^2]$$

$$= \gamma/2\ \left[\Delta\vec{\mathbf{r}}_1\Delta\vec{\mathbf{r}}_2\Delta\vec{\mathbf{r}}_3\right]\Big(\underbrace{\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}}_{\text{the connectivity matrix, }\boldsymbol{\Gamma}} \otimes \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\Big)\begin{bmatrix}\Delta\vec{\mathbf{r}}_1\\ \Delta\vec{\mathbf{r}}_2\\ \Delta\vec{\mathbf{r}}_3\end{bmatrix}$$

Figure S1 (a) The relative orientations of vectors $\vec{\mathbf{r}}_{ij}$, $\vec{\mathbf{r}}_{ij}^0$, $\Delta\vec{\mathbf{r}}_j$ and $\Delta\vec{\mathbf{r}}_i$ **(b)** The elastic potential of a tri-peptide molecule, represented by C_α atoms only. The cutoff R_c here is set small enough for the atom pairs 1-2 and 2-3 in contact but not for 1-3. Provided a certain pair ij ($i \neq j$) is in contact (their linear separation < R_c), the off-diagonal element ij of the connectivity matrix $\boldsymbol{\Gamma}$ is set to be -1; otherwise 0. The diagonal elements are the negative sums of the off-diagonal elements in each individual row (or column; since $\boldsymbol{\Gamma}$ is symmetric). Here $\left[\Delta\vec{\mathbf{r}}_1\Delta\vec{\mathbf{r}}_2\Delta\vec{\mathbf{r}}_3\right] = \left[\Delta x_1\Delta y_1\Delta z_1\Delta x_2\Delta y_2\Delta z_2\Delta x_3\Delta y_3\Delta z_3\right]$

II To derive the eGNM Hessian, $\mathbf{\Gamma}\otimes\mathbf{I}$ matrix, from GNM potential

$$E_{GNM} = \sum_{i,j=1}^{N}\frac{\gamma}{2}(\Delta\vec{\mathbf{r}}_j - \Delta\vec{\mathbf{r}}_i)\bullet(\Delta\vec{\mathbf{r}}_j - \Delta\vec{\mathbf{r}}_i)H(R_c - |\vec{\mathbf{r}}_{ij}^0|)$$
$$= \sum_{i,j=1}^{N}\frac{\gamma}{2}[(\Delta x_j - \Delta x_i)^2 + (\Delta y_j - \Delta y_i)^2 + (\Delta z_j - \Delta z_i)^2]H(R_c - |\vec{\mathbf{r}}_{ij}^0|) \qquad \text{(SI-11)}$$

For those j that contact with i,

$\frac{\partial E}{\partial \Delta x_i} = -\gamma\sum_{j=1}^{N}(\Delta x_j - \Delta x_i)$; $\frac{\partial^2 E}{\partial \Delta x_i^{\,2}} = \gamma\sum_{j=1}^{N} = \gamma n_i = \frac{\partial^2 E}{\partial \Delta y_i^{\,2}} = \frac{\partial^2 E}{\partial \Delta z_i^{\,2}}$; n_i is the number of neighbors of i within a given contact distance R_c.

$\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta y_i} = 0$; $\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta z_i} = 0$;

Hence, the diagonal superelement ii in $\mathbf{\Gamma}\otimes\mathbf{I}$ comes with the form

$$\begin{bmatrix} \gamma n_i & 0 & 0 \\ 0 & \gamma n_i & 0 \\ 0 & 0 & \gamma n_i \end{bmatrix}$$

For off-diagonal superelement ij in $\mathbf{\Gamma}\otimes\mathbf{I}$,

$\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta x_j} = -\gamma$; $\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta y_j} = 0$; $\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta z_j} = 0$,

hence of the form $\begin{bmatrix} -\gamma & 0 & 0 \\ 0 & -\gamma & 0 \\ 0 & 0 & -\gamma \end{bmatrix}$

As a result, $\mathbf{\Gamma}\otimes\mathbf{I}$ comprises elements such as

$$\begin{bmatrix}
\gamma n_1 & 0 & 0 & -\gamma & 0 & 0 & -\gamma & 0 & 0 & 0 & & \\
0 & \gamma n_1 & 0 & 0 & -\gamma & 0 & 0 & -\gamma & 0 & 0 & \cdots & \cdots \\
0 & 0 & \gamma n_1 & 0 & 0 & -\gamma & 0 & 0 & -\gamma & 0 & & \\
-\gamma & 0 & 0 & \ddots & & & & \vdots & & & & \\
0 & -\gamma & 0 & & \ddots & & & \vdots & & & & \\
0 & 0 & -\gamma & & & \ddots & & \vdots & & & & \\
-\gamma & 0 & 0 & & & & \gamma n_i & 0 & 0 & & & \\
0 & -\gamma & 0 & \cdots & \cdots & \cdots & 0 & \gamma n_i & 0 & & & \\
0 & 0 & -\gamma & & & & 0 & 0 & \gamma n_i & & & \\
0 & 0 & 0 & & & & & & & \ddots & & \\
 & \vdots & & & & & & & & & \ddots & \\
 & \vdots & & & & & & & & & & \ddots
\end{bmatrix}_{3N\times 3N}$$

The fact that GNM modes contain rigid-body rotation was not widely conceived before. One of the reasons is that it is hard to examine rotation in the *N*-dimensional space instead of 3*N*. eGNM concept provides a correct scaffold and resolves modes in 3*N*-dimension hence an easier confirmation and characterization of external rotation in modes.

III Derivation of the Hessian, H, based on ANM potential

$$E_{ANM} = \sum_{i,j=1}^{N} \frac{\gamma}{2}(L_{ij} - L_{ij}^0)^2 H(R_c - |L_{ij}^0|) = \frac{\gamma}{2}\Delta\mathbf{R}^T\mathbf{H}\Delta\mathbf{R} \qquad \text{(SI-12)}$$

where $L_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}$

$L^0{}_{ij} = \sqrt{(x_j^o - x_i^o)^2 + (y_j^o - y_i^o)^2 + (z_j^o - z_i^o)^2}$

and here $x_i = x_i^o + \Delta x_i$

so that $L_{ij} = \sqrt{(\Delta x_j - \Delta x_i + x_j^o - x_i^o)^2 + (\Delta y_j - \Delta y_i + y_j^o - y_i^o)^2 + (\Delta z_j - \Delta z_i + z_j^o - z_i^o)^2}$

For those *j* that contact with *i*, we take the first derivative of E_{ANM} (thereafter denoted as *E*) against Δx_i to obtain

$$\mathrm{F}_i^{\mathrm{x}} = \frac{\partial E}{\partial \Delta x_i} = \gamma \sum_j (L_{ij} - L_{ij}^0)\frac{\partial L_{ij}}{\partial \Delta x_i} = \gamma \sum_j (L_{ij} - L_{ij}^0)\frac{-(\Delta x_j - \Delta x_i + x_j^o - x_i^o)}{L_{ij}} \qquad \text{(SI-13)}$$

Take the derivative of the above obtained again to obtain the second derivative of *E*

$$\frac{\partial^2 E}{\partial \Delta x_i^{\,2}} = \gamma \sum_{j=1}^{N} \frac{(\Delta x_j - \Delta x_i + x_j^o - x_i^o)^2}{L_{ij}^2} + (L_{ij} - L_{ij}^0)\frac{1}{L_{ij}} + (L_{ij} - L_{ij}^0)\frac{-(\Delta x_j - \Delta x_i + x_j^o - x_i^o)^2}{L_{ij}^3}$$

When the second derivative is taken at the equilibrium, $L_{ij} = L_{ij}^0$, the above expression is reduced to

$$\left.\frac{\partial^2 E}{\partial \Delta x_i^{\,2}}\right|_{\Delta x_i, \Delta y_i, \Delta z_i = 0} = \gamma \sum_{j=1}^{N} \frac{(x_j^o - x_i^o)^2}{(L_{ij}^0)^2} \qquad \text{(SI-14)}$$

Similarly, $\left.\frac{\partial^2 E}{\partial \Delta y_i^{\,2}}\right|_{\Delta x_i, \Delta y_i, \Delta z_i = 0} = \gamma \sum_{j=1}^{N} \frac{(y_j^o - y_i^o)^2}{(L_{ij}^0)^2}$; $\left.\frac{\partial^2 E}{\partial \Delta z_i^{\,2}}\right|_{\Delta x_i, \Delta y_i, \Delta z_i = 0} = \gamma \sum_{j=1}^{N} \frac{(z_j^o - z_i^o)^2}{(L_{ij}^0)^2}$

and $$\left.\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta y_i}\right|_{\Delta x_i, \Delta y_i, \Delta z_i = 0} = \gamma \sum_{j=1}^{N} \frac{(y_j^o - y_i^o)(x_j^o - x_i^o)}{(L_{ij}^0)^2} \qquad \text{(SI-15)}$$

Similar approach extended when *i* is not equal to *j*,

$$\left.\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta x_j}\right|_{\Delta x_i, \Delta y_i, \Delta z_i = 0} = -\gamma \frac{(x_j^o - x_i^o)^2}{(L_{ij}^0)^2} \qquad \text{(SI-16)}$$

$$\left.\frac{\partial^2 E}{\partial \Delta x_i \partial \Delta y_i}\right|_{\Delta x_i, \Delta y_i, \Delta z_i = 0} = -\gamma \frac{(y_j^o - y_i^o)(x_j^o - x_i^o)}{(L_{ij}^0)^2} \qquad \text{(SI-17)}$$

The 3N×3N Hessian **H** can therefore be constructed. Also, it can be derived similarly as described in the Section I that

$$<\Delta \mathbf{R} \Delta \mathbf{R}^{\boldsymbol{T}}> = \frac{k_B T}{\gamma} \mathbf{H}^{-1}. \qquad \text{(SI-18)}$$

The diagonalization of Hessian gives six vanishing eigenvalues such that

$$\mathbf{H} = \sum_{k=7}^{3N} \frac{1}{\lambda''_k} \mathbf{V}''_{\mathbf{k}} \mathbf{V}''^{\mathbf{T}}_{\mathbf{k}} \quad \text{(see also Ref (4)).} \qquad \text{(SI-19)}$$

IV Three trivial modes obtained from diagonalization of Hessian when the model is translational invariant

In Cartesian space, for any model that is translational invariant (such as non-condensed GNM), 3 zero eigen modes are generated in the diagonalization of Hessian, and another 3 zero modes are generated if it is also rotational invariant (such as ENM). A simple explanation is as follows, which justifies why eGNM Hessian, $\mathbf{\Gamma} \otimes \mathbf{I}$, results in 3 zero eigen modes for the external translation.

Let us consider a one-dimension system along which N degrees of freedom are located, and the overall energy for this system is $E(x_1, x_2, .. x_N)$ which is a function of positions of all the d.o.f. When the system is translational invariant in energy,

$$E(x_1, x_2, .. x_N) = \quad E(x_1+c, x_2+c, .. x_N+c) \qquad \text{(SI-20)}$$

where c is the amount of displacement by which all the d.o.f systematically translated. The E is not changed with c to be translational invariant. The first derivative of E against c should therefore be zero. That is

$$\frac{dE}{dc} = 0 = \sum_{j=1}^{N} \frac{\partial E}{\partial x_j} \frac{dx_j}{dc} \qquad \text{(SI-21)}$$

We also note that the term $\frac{dx_j}{dc}$ is unity when the system is in rigid-body translation (x_j has to change as much as c does), which leads to $\sum_{j=1}^{N} \frac{\partial E}{\partial x_j} = 0$

We take partial derivative against x_i on both sides of the above expression to obtain

$$\sum_{j=1}^{N} \frac{\partial E}{\partial x_i \partial x_j} = 0 \qquad \text{(SI-22)}$$

$\frac{\partial E}{\partial x_i \partial x_j}$ are the Hessian elements in those simple models discussed in the study. According to the derived equality,

$$\mathbf{H}\,\mathbf{v} = \lambda\,\mathbf{v} = 0 \qquad \text{(SI-23)}$$

where $\mathbf{v}$ is the normalized N-d column vector for the external translation of the system such that

$$\mathbf{v} = [\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{N}} \cdots \frac{1}{\sqrt{N}}]^{\mathrm{T}}. \qquad \text{(SI-24)}$$

Hence, it becomes clear that translational invariance exemplified in the one-dimension system results in one zero eigen mode. In the Cartesian space, we therefore obtain three trivial modes due to the translational invariance. The trivial modes that result from the rotational invariance of the potential (such as ANM) can be rationalized in a similar way.

V Force and Torque experienced in GNM- and ANM-defined Potential

GNM - Energy function defined by vectorial differences of atomic displacements

According to Eq. SI-1, force acting on atom *I* is

$$\mathbf{f}_i = -\frac{\partial E_{GNM}}{\partial \mathbf{r}_i} = \sum_{j\neq i}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)(\mathbf{r}_{ij} - \mathbf{r}_{ij}^0) \qquad \text{(SI-25)}$$

where $\mathbf{f}_i = [\, f_{x,i} \quad f_{y,i} \quad f_{z,i}\,]^T$ and $\frac{\partial}{\partial \mathbf{r}_i} = [\frac{\partial}{\partial x_i} \quad \frac{\partial}{\partial y_i} \quad \frac{\partial}{\partial z_i}]^T$

Total force acting on the molecule.

$$\mathbf{F}_{total} = \sum_{i=1}^{N} \mathbf{f}_i = \sum_{j\neq i}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)(\mathbf{r}_{ij} - \mathbf{r}_{ij}^0) = \mathbf{0} \qquad \text{(SI-26)}$$

To derive Eq. SI-26 we use $\mathbf{r}_{ij} = -\mathbf{r}_{ji}$ and $\mathbf{r}_{ij}^0 = -\mathbf{r}_{ji}^0$.

Torque acting on the molecule;

$$\begin{aligned}
\boldsymbol{\tau} &= \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{f}_i \\
&= \sum_{i=1}^{N} \mathbf{r}_i \times \sum_{j\neq i}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)(\mathbf{r}_{ij} - \mathbf{r}_{ij}^0) \\
&= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)\left(\mathbf{r}_i \times (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0) + \mathbf{r}_j \times (\mathbf{r}_{ji} - \mathbf{r}_{ji}^0)\right) \\
&= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)\left(\mathbf{r}_i \times (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0) - \mathbf{r}_j \times (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0)\right) \\
&= -\sum_{i=1}^{N} \sum_{j=i+1}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)\left(\mathbf{r}_{ij} \times (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0)\right) \\
&= -\sum_{i=1}^{N} \sum_{j=i+1}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)\left(\mathbf{r}_{ij}^0 \times \mathbf{r}_{ij}\right) \\
&= -\sum_{i=1}^{N} \sum_{j=i+1}^{N} \gamma H(R_c - |\mathbf{r}_{ij}^0|)\left(\mathbf{r}_{ij}^0 \times \Delta\mathbf{r}_{ij}\right) \qquad \text{(SI-27)}
\end{aligned}$$

To derive Eq. SI-27, $\mathbf{r}_{ij} = -\mathbf{r}_{ji}$, $\mathbf{r}_{ij} \times \mathbf{r}_{ij} = \mathbf{0}$ and $\Delta\mathbf{r}_{ij} = \mathbf{r}_{ij} - \mathbf{r}_{ij}^0$ are used. We note that $\boldsymbol{\tau} = \mathbf{0}$ when $\Delta\mathbf{r}_{ij} \,//\, \mathbf{r}_{ij}^0$ for all $i, j \in |\mathbf{r}_{ij}^0| < R_c$. Non-vanishing torque acts on the molecule when molecule is subject to a rigid-body rotation. Hence, GNM has translational but not rotational invariance.

ANM - Energy function defined by the difference of inter-atomic distances

We know that ANM potential is the sum of a function ϕ that is a function of r_{ij} such that

$$E_{ANM} = \sum_{i,j=1,j>i}^{N} \phi(r_{ij}) \tag{SI-28}$$

where, $r_{ij} = \left|\mathbf{r}_{ij}\right|$. The force acting on atom *I* is

$$\mathbf{f}_i = -\frac{\partial E_{ENM}}{\partial \mathbf{r}_i} = \sum_{j>i}^{N} \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} \tag{SI-29}$$

Total force acting on the molecule;

$$\mathbf{F}_{total} = \sum_{i=1}^{N} \mathbf{f}_i = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} = \mathbf{0} \tag{SI-30}$$

To derive Eq. SI-30, we use $\mathbf{r}_{ij} = -\mathbf{r}_{ji}$ and $r_{ij} = r_{ji}$.

Torque acting on the molecule;

$$\begin{aligned}
\boldsymbol{\tau} &= \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{f}_i \\
&= \sum_{i=1}^{N} \mathbf{r}_i \times \sum_{j>i}^{N} \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} \\
&= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\mathbf{r}_i \times \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} + \mathbf{r}_j \times \frac{\partial \phi(r_{ji})}{\partial r_{ji}} \frac{\mathbf{r}_{ji}}{r_{ji}} \right) \\
&= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\mathbf{r}_i \times \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} - \mathbf{r}_j \times \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \frac{\mathbf{r}_{ij}}{r_{ij}} \right) \\
&= -\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\frac{1}{r_{ij}} \frac{\partial \phi(r_{ij})}{\partial r_{ij}} \mathbf{r}_{ij} \times \mathbf{r}_{ij} \right) = \mathbf{0}
\end{aligned} \tag{SI-31}$$

To derive Eq. SI-31, the relation $\mathbf{r}_{ij} = -\mathbf{r}_{ji}$ and $r_{ij} = r_{ji}$ are used. Thus there is neither total force nor torque that acts on a molecule for an energy function defined as Eq. SI-28. Hence, ANM has both translational and rotational invariance.

VI Normal mode refinement study by Kidera et al sheds light on the contributions of internal and external motions to B-factors

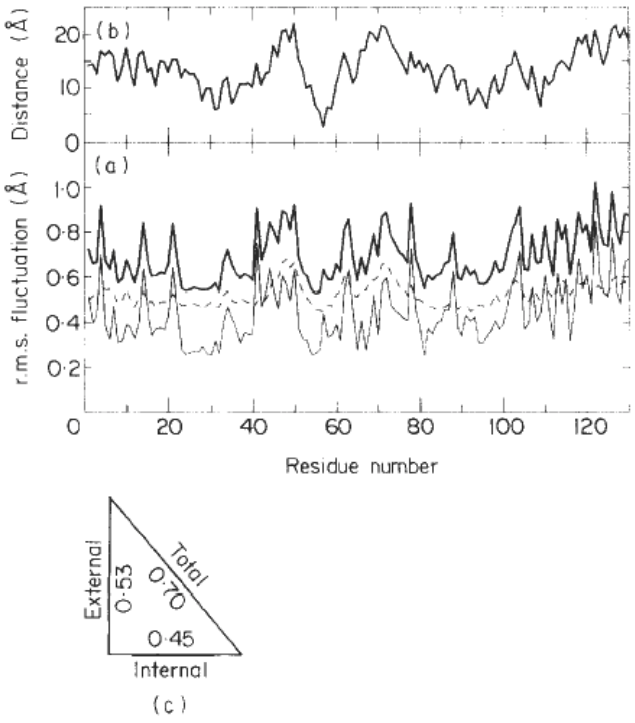


Figure 1. (a) r.m.s. fluctuations of non-hydrogen atoms in human lysozyme determined by the normal mode refinement using the model of D100F43E and averaged within each residue. The heavy, broken and thin curves are for the total, external and internal fluctuations, respectively. (b) Distance between centers of gravity of non-hydrogen atoms in the molecules and in each residue. This profile has the following correlation coefficients for the curves in (a): 0·63, 0·81 and 0·51, for the total, external and internal fluctuations, respectively. (c) The average amplitudes of the total, external and internal fluctuations are shown schematically in a right-angled triangle. Squares of the latter 2 equal the square of the former.

Figure S2 The figure is reproduced from Akinori Kidera, Koji Inaka, Masaaki Matsushima and Nobuhiro Gō (1992) "Normal Mode Refinement: Crystallographic Refinement of Protein Dynamic Structure II. Application to Human Lysozyme" *J. Mol. Biol.* 225, 477-486

The panel **(a)** of **Fig. S2** demonstrates three types of motions – total, internal and external. In this normal mode refinement study, the contribution of internal vibrations to B-factors are modeled by all-atom normal mode analysis (NMA) and contribution of external motions to B-factors are modeled by TLS model (21). Two facts are to be noted here.

First, indeed, external motions have a dominant contribution to the size of the total fluctuation, 0.53/0.70, when internal motions take the share of 0.45/0.70;

Secondly, despite of larger contribution of external motions in its size, *the shape of the profile of the total fluctuation are dominantly determined by that of internal motions.*

External motions on the other hand have a relatively flat profile (featureless). Abscissa is the residue index; Ordinate is the r.m.s fluctuation (Å). We know that B-factors are the size of atom fluctuations (the total fluctuations) multiplied by a constant. We also know that the fluctuation profiles of residue internal motions can be predicted by ANM and EPIRM. Pearson correlation between these two profiles (B-factors and internal motions) indicates given models' applicability to describe B-factors. It is mathematically trivial but worth noting – the correlation coefficient does not change with the absolute size of the two profiles. The correlation is high as long as the shapes of the two profiles are similar. This mathematical judgment has been used in most of (if not all) the published studies.

VII Further remarks on EPIRM

Understand the difference between ANM modes and eGNM modes

Derived from a potential that is translation/rotation invariant, ANM modes contain no contributions to the translational momentum and angular momentum of the system. Take a tripeptide system for example (Fig S3). A given ANM internal mode $\mathbf{U_A}$ has nine components in the Cartesian coordinate and $\mathbf{U_A} = [\mathbf{V}_1^{\mathrm{T}} \ \mathbf{V}_2^{\mathrm{T}} \ \mathbf{V}_3^{\mathrm{T}}]^{\mathrm{T}}$; Each $\mathbf{V}_k$ is a 3-d column vector indicating the size and directions of one residue of the three; 'T' is transpose. Let us assume the simplest case where the mass center of the system does not move and the system does not rotate (however the following discussions also hold for systems translate in constant velocity and/or rotate in constant speed). The following two conditions have to be held.

(1) $\sum_{k=1}^{3} m_k \mathbf{V}_k / \sum_{k=1}^{3} m_k = 0$, m_k is the mass of residue k. Let us assume equal mass for the residues such that $m_1 = m_2 = m_3 = m$

(2) $\mathbf{L} = \sum_{k=1}^{3} \mathbf{r}_k \times m\mathbf{V}_k = 0$; $\mathbf{r}_k$ is the position vectors for residue k; '×' is the cross product.

Rigorously speaking, every $\mathbf{V}_k$ in the aforementioned conditions should be replaced by $c\mathbf{V}_k/\alpha$ where $c=(k_BT/\gamma)^{1/2}$ and $\alpha=2\pi(1/\gamma\lambda_k)^{1/2}$ in order to have $\mathbf{V}_k$ in the unit of velocity; otherwise $\mathbf{V}_k$ is dimensionless.

Every ANM internal mode fulfills these two conditions; On the other hand, non-trivial eGNM modes (led by nonzero eigenvalues) fulfill the first condition but not the second.

How is a EPIRM mode derived from a eGNM mode?

This nonzero L is also known equal to $\mathbf{I\omega}$. $\mathbf{I}$ is the 3×3 moment of inertia tensor.

$$\mathbf{I}=\begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} \text{ where } I_{xx}=\sum_{k=1}^{3} m_k(y_k^2+z_k^2);\ I_{yy}=\sum_{k=1}^{3} m_k(x_k^2+z_k^2);\ I_{zz}=\sum_{k=1}^{3} m_k(x_k^2+y_k^2);$$

$$I_{xy}=I_{yx}=-\sum_{k=1}^{3} m_k x_k y_k\ ;\ I_{yz}=I_{zy}=-\sum_{k=1}^{3} m_k y_k z_k\ ;\ I_{xz}=I_{zx}=-\sum_{k=1}^{3} m_k x_k z_k$$

The angular velocity can be calculated as $\boldsymbol{\omega}=\mathbf{I}^{-1}\mathbf{L}$; the direction of vector $\boldsymbol{\omega}$ points to the axis that the whole system rotates about. As a result, one way to remove such rigid rotation is to remove the contribution to this rotation from each residue. In another word, we should remove the component in $\mathbf{V}_k$ that directly contributes to $\boldsymbol{\omega}$. This contribution is known to be $\boldsymbol{\omega}\times\mathbf{r}_k$. Hence, $\mathbf{V}_k$ - $\boldsymbol{\omega}\times\mathbf{r}_k$ would result in a new vector $\mathbf{V}_k$' that does not contribute to the rigid-body rotation. Reassembling the $\mathbf{V}_k$' (k =1..3) returns us the EPIRM mode.

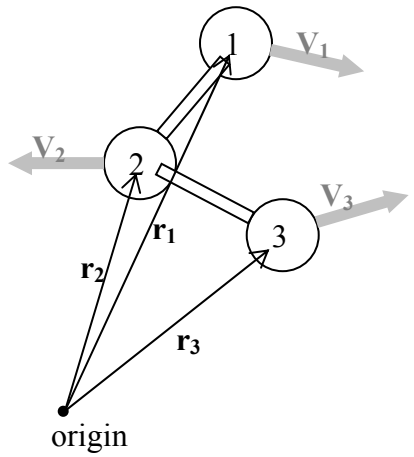


Figure S3 A tripeptide system motions described by a given normal mode

VIII ANISOTROPY predictions by EPIRM and ANM

In high-resolution PDB structure files, there are 6 anisotropic displacement parameters (ADPs) associated with each atom. These ADPs are three diagonal and three off-diagonal elements of a 3×3 covariance matrix **U** associated with the Gaussian probability distribution of atomic fluctuations in space (14).

The symmetric matrix $\mathbf{U}_i$ for atom i = $\begin{bmatrix} U_{i,xx} & U_{i,xy} & U_{i,xz} \\ U_{i,yx} & U_{i,yy} & U_{i,yz} \\ U_{i,zx} & U_{i,yz} & U_{i,zz} \end{bmatrix}$ where U_{xx}, U_{yy} and U_{zz} are the ADPs forming the diagonal elements while U_{xy}, U_{yz} and U_{zx} form the off-diagonal elements of $\mathbf{U}_i$.

Diagonalizing $\mathbf{U}_i$ renders three eigenvectors and three eigenvalues. The eigenvectors give three principal axes of an ellipsoid and the eigenvalues give the spread of the ellipsoid on these axes. ISOTROPY is defined as the minimal eigenvalue divided by the maximal eigenvalue. A unity value would mean perfect isotropic fluctuations while values close to zero suggest highly anisotropic fluctuations of the atom.

ADP-derived ISOTROPY for all the C_α atoms in the proteins listed in Table S1 are computed and compared with the ISOTROPY derived from the covariance matrices of EPIRM (Eq. 4) and ANM (Eq. SI-18) for the same C_α atoms. We find an average correlation of 0.3144±0.031 between experiment and EPIRM, and that of 0.3306±0.028 between experiment and ANM. The two correlations are statistically identical. The low correlations are consistent with our previous finding (14).

Care should be taken that even perfect predictions for ISOTROPY do not necessarily suggest any good prediction for directionalities and sizes of atom fluctuations. This is self-evident from the definition of ISOTROPY.

GNM is an isotropic model and ANM is anisotropic. One may intuitively think that ANM would perform less well than GNM in predicting atom dynamics that are highly isotropic. We examine such intuition using the protein set in Table S1. We define an AVERAGE ISOTROPY for a protein which is the average of the ISOTROPY values of all the C_αs in that protein. On the other hand, the correlation between B-factors and ANM/GNM predictions for this protein is calculated. We plot the quantity (AVERAGE ISOTROPY, correlation) for each of the proteins in the figure below (**Fig S4**).

The result shows both GNM and ANM better predict B-factors when proteins have higher dynamics isotropy (**Fig S4**). The ISOPTROPY-dependent correlation increase demonstrates a 0.41 correlation for both GNM and ANM (**Fig S4**). This slightly counter-intuitive result indicates that ENMs generally better predicts isotropic data (e.g. sizes of the fluctuations) than anisotropic data (e.g. directions of the fluctuations) despite of their inherited assumptions, namely, isotropic or anisotropic. This is also noted by other studies (13, 14, 16).

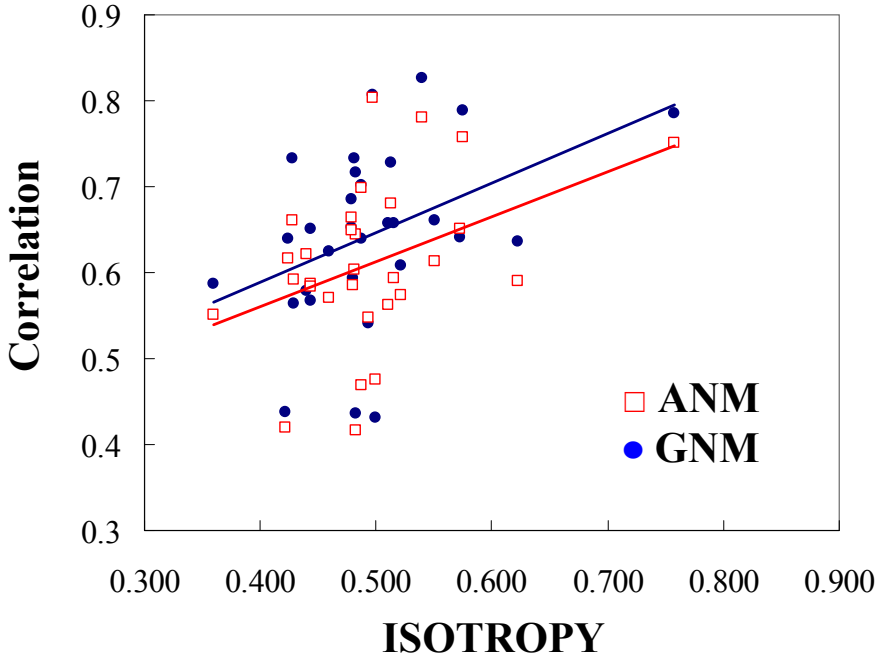


Figure S4 B-factor predictability (Correlation) versus dynamics ISOTROPY plotted for 30 high-resolution structures

IX Catalytic residues of Ricin Hydrolase distributed near the (external) rotation axes identified from the first two eGNM modes

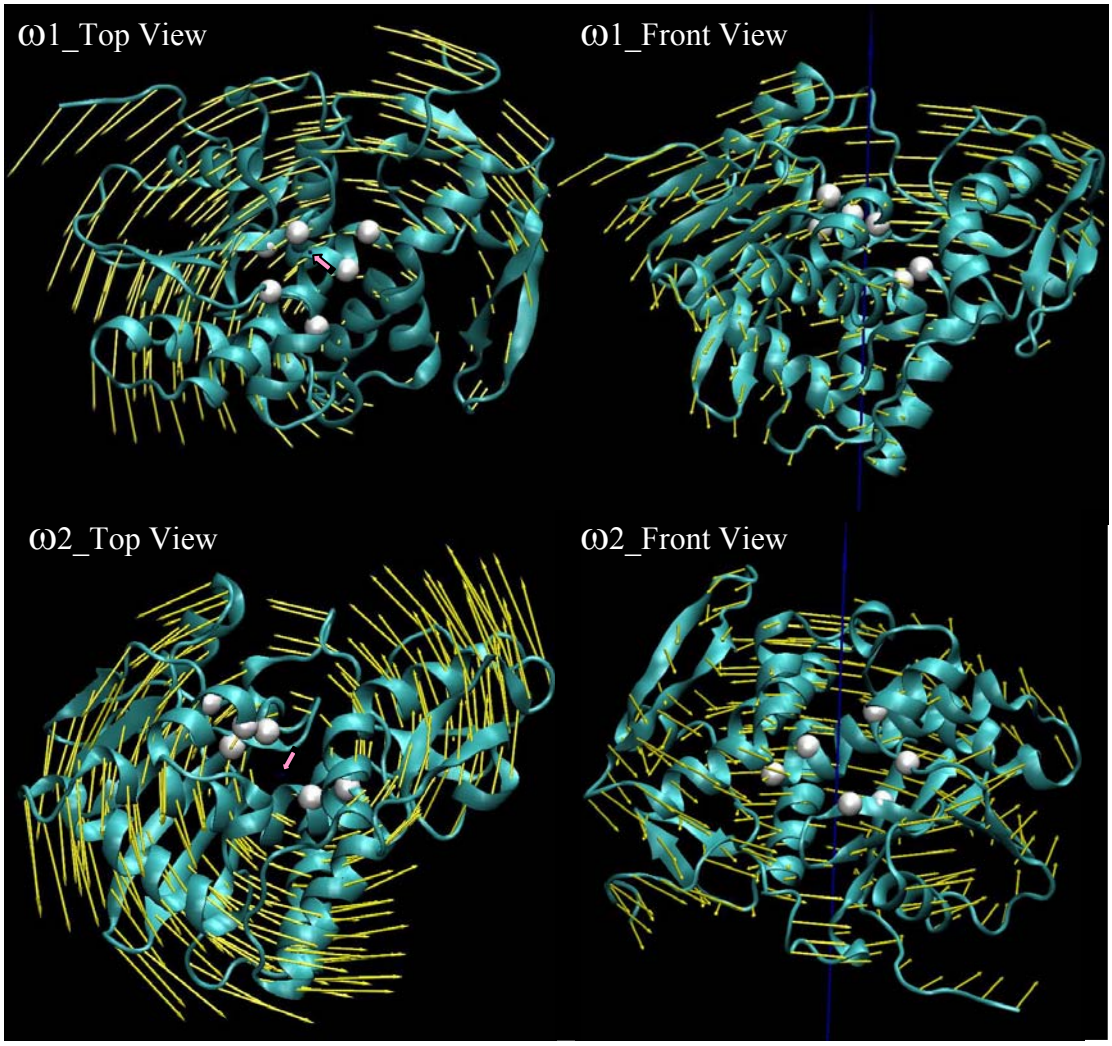


Figure S5 Distribution of catalytic residues of Ricin Hydrolase near the rotation axes
ω1 and **ω**2 are the angular velocity vectors indicating the size and direction of the external rotation contaminated in the eGNM mode 1 and 2, respectively. The blue arrows, going through the mean locations of the proteins, point toward the directions of **ω**1 and **ω**2, but their exaggerated sizes do not correspond to the real magnitude of **ω**1 and **ω**2. The yellow arrow pointing from the *i*-th residue indicates the vector $(1/c)\boldsymbol{\omega}_k \times \mathbf{r}_i$ (see Eq. 6 in the main text; k= 1 or 2 herein) that is the componential contribution of a certain residue's movement to the rigid-body rotation, or specifically, the **ω**. Note that the length of the arrows are drawn in proportion to the magnitude of $(1/c)\boldsymbol{\omega}_k \times \mathbf{r}$ but exaggerated for easy visualization. The C_α atoms of the catalytic residues Y80, V81, G121, Y123, E177 and R180 are shown in white spheres. In the top views, the small pink arrows indicate the tips of the rotation axes.