

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/275519454>

Exploring Free Energy Landscapes of Large Conformational Changes: Molecular Dynamics with Excited Normal Modes

ARTICLE *in* JOURNAL OF CHEMICAL THEORY AND COMPUTATION · APRIL 2015

Impact Factor: 5.5 · DOI: 10.1021/acs.jctc.5b00003

CITATIONS

2

READS

154

4 AUTHORS, INCLUDING:



Mauricio Garcia de Souza Costa

Fundaão Oswaldo Cruz

15 PUBLICATIONS 67 CITATIONS

[SEE PROFILE](#)



Paulo Ricardo Batista

Fundaão Oswaldo Cruz

16 PUBLICATIONS 155 CITATIONS

[SEE PROFILE](#)



David Perahia

French National Centre for Scientific Research

134 PUBLICATIONS 3,353 CITATIONS

[SEE PROFILE](#)

Article

Exploring free energy landscapes of large conformational changes: Molecular Dynamics with Excited Normal modes

Mauricio Garcia de Souza Costa, Paulo Ricardo Batista, Paulo Mascarello Bisch, and David Perahia

J. Chem. Theory Comput., Just Accepted Manuscript • Publication Date (Web): 20 Apr 2015

Downloaded from <http://pubs.acs.org> on April 27, 2015

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.



ACS Publications

High quality. High impact.

Journal of Chemical Theory and Computation is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society.

However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

1
2
3 **Exploring free energy landscapes of large conformational changes:**
4 **Molecular Dynamics with Excited Normal modes**
5
6
7
8
9

10 Mauricio G.S Costa^{1,2,3}, Paulo R Batista¹, Paulo M Bisch² and David Perahia^{3,§}
11
12

13 ¹ Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro - RJ, Brazil.
14
15

16 ² Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, 21949-901,
17 Rio de Janeiro, Brazil.
18
19

20 ³ Laboratoire de Biologie et de Pharmacologie Appliquée, Ecole Normale Supérieure de Cachan,
21 Centre National de la Recherche Scientifique, 61, F-94235 Cachan, France
22
23
24
25
26

27 Email addresses:
28
29
30
31
32

33 MGSC: mauriciocosta@fiocruz.br
34 PRB: pbatista@fiocruz.br
35 PMB: pmbisch@biof.ufrj.br
36 DP: david.perahia@ens-cachan.fr
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

§To whom correspondence should be addressed. E-mail: david.perahia@ens-cachan.fr

Abstract

Proteins are found in solution as ensembles of conformations in dynamic equilibrium. Exploration of functional motions occurring on μ s-ms time scales by Molecular Dynamics (MD) simulations still remains computationally challenging. Alternatively, normal mode (NM) analysis is a well suited method to characterize intrinsic slow collective motions, often associated to protein function, but the absence of anharmonic effects preclude a proper characterization of conformational distributions in a multidimensional NM space. Using both methods jointly appears as an attractive approach that allows an extended sampling of the conformational space. In line with this view the MDeNM (Molecular Dynamics with excited Normal Modes) method presented here consists in multiple-replicas short MD simulations in which motions described by a given subset of low frequency NMs are kinetically excited. This is achieved by adding additional atomic velocities along several randomly determined linear combinations of NM vectors, thus allowing an efficient coupling between slow and fast motions. The relatively high-energy conformations generated with MDeNM are further relaxed with standard MD simulations enabling to determine free energy landscapes. Two widely studied proteins were selected as examples: hen egg lysozyme and HIV-1 protease. In both cases, MDeNM provides a larger extent of sampling in a few nanoseconds, outperforming long standard MD simulations. It was noticed a high degree of correlation with motions inferred from experimental sources (X-ray, EPR and NMR), and with free energy estimations obtained by metadynamics. Finally, the large sets of conformations obtained with MDeNM can be used in a better characterization of relevant dynamical populations, allowing a better interpretation of experimental data such as SAXS and NMR spectra.

1. Introduction

Proteins are dynamic entities constantly adopting distinct conformations. The associated free energy landscapes (FEL) are extremely rugged, containing a large number of similar microstates, with equivalent energy levels separated by low energy barriers (multiple minima landscapes)¹. Recent NMR studies on proteins have revealed the existence of sparsely populated conformations (metastable states) but their transient nature prevents their structures to be experimentally solved²⁻³. Accompanying the equilibrium between multiple states, a rich hierarchy of protein movements takes place occurring in distinct time scales. They may range from small individual atomic displacements to large amplitude collective motions⁴⁻⁵. These latter slow motions are often linked to molecular mechanisms associated to signaling, allostery and ligand binding⁶. Thus, a deep understanding of FELs associated to collective motions is required to characterize functional mechanisms and metastable states not easily accessible from experimental measures.

One of the challenges in computational biology is to structurally characterize functional motions occurring at large time scales. The standard all-atom molecular dynamics with the current available computational resources are still insufficient to fully explore slow functional motions^{7,8}. Moreover, multiple replicas are required to ensure the statistical relevance and convergence of the results. Besides, normal modes (NM) analysis is well-suited method to explore slow motions, and has gained increased popularity over the last decades^{5,9}. Indeed, there are many examples demonstrating how the directions of lower frequency NMs often present significant overlap with experimentally described conformational changes⁵.

NM analysis is based on a quadratic approximation of the energy surface (ES) around a given minimum-energy structure, and consequently does not take into account the anharmonic aspects of ES¹⁰⁻¹¹. On the other hand, MD systematically fails to cover the entire range of large amplitude conformational changes due to time scale limitations⁵. Alternatively, to overcome some of the limitations of standard simulation methods as mentioned before, several hybrid approaches have been developed to better explore the protein conformational ES¹². Generally, the aim is to accelerate sampling during the simulation (*e.g.* change of the potential energy function, use of a reaction coordinate, etc.) allowing an efficient exploration of the protein PES¹³. Among the enhanced sampling simulations, collective variables (CV) based methods, such as umbrella sampling and metadynamics are often used¹⁴⁻¹⁵. The former requires lengthy simulations under equilibrium conditions through successive windows, using harmonic potentials along a reaction coordinate. On the other hand, in metadynamics simulations, successive repulsive gaussians are added to the potential to avoid revisiting of sampled states, allowing escaping from minima. However, this method is highly dependent of a careful choice of parameters, demanding extensive validation¹⁵. Further, both methods are not effective when using more than two CVs.

Recently, there has been some attempts to explore low frequency NM directions (or their linear combinations) using hybrid methods based on energy minimizations or MD simulations¹⁶⁻¹⁸. Nonetheless, the high number of degrees of freedom associated to an increased multidimensional NM space leads to a combinatorial problem. Consequently, only few (up to three) relevant modes are normally investigated. They should be chosen based on their prior ‘known’ relevance to describe a given phenomenon. Neglecting other

1
2
3 important modes could result in an underestimation of entropic contributions of the other
4 overall motions¹⁹.
5
6

7 Here, in order to circumvent most of the limitations mentioned above, an approach
8 termed as MDeNM (**M**olecular **D**ynamics with **e**cited **N**ormal **M**odes) is presented,
9 where the space defined by a set of low-frequency NMs is extensively explored. The
10 underlying principle of MDeNM is the kinetic excitation of low frequency modes during
11 several short MD simulations, therefore coupling local and global motions efficiently.
12
13 Our approach is based on a multi-replica scheme in which a large number of MD
14 simulations are performed, each of them corresponding to the excitation of a randomly
15 determined linear combination of selected normal modes. Unbiased FELs in the NM
16 space can be further obtained by carrying out standard MD simulations starting from
17 representative uniformly distributed excited conformations previously generated.
18
19

20 Two widely studied proteins, the hen-egg lysozyme (HEWL) and HIV-1 protease (PR)
21 were considered as test cases. We demonstrate here from multiple analyses how short
22 MDeNM simulations outperformed standard MD simulations providing a larger extent of
23 sampling. Moreover, better correlations with motions described by experiments (X-ray,
24 NMR and EPR) were also obtained. Lastly, FELs obtained with MDeNM simulations
25 were in agreement with metadynamics calculations.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2. Theory

NM is well suited to describe collective motions not easily explored in standard MD simulations. The purpose of the MDeNM approach is the kinetic excitation of such movements during MD simulations. To this end, additional atomic velocities are assigned along linear combinations of NM vectors yielding an incremental temperature (ΔT_{nm}) that adds to the current temperature of an equilibrated system (T_{curr}). In a constant temperature MD simulation coupled to a heat bath, this incremental NM temperature is dissipated during the course of the simulation, which may require multiple re-excitations of the collective motions in order to reach larger displacements, especially in proteins presenting complex architectures.

2.1. MDeNM approach

Figure. 1 depicts the scheme of the MDeNM excitation procedure. This approach was designed as a CHARMM script to be run with this program²⁰. Additional details of the multi-excitation scheme and a conformational exploration protocol designed to the calculation of FELs are given in Supporting Information.

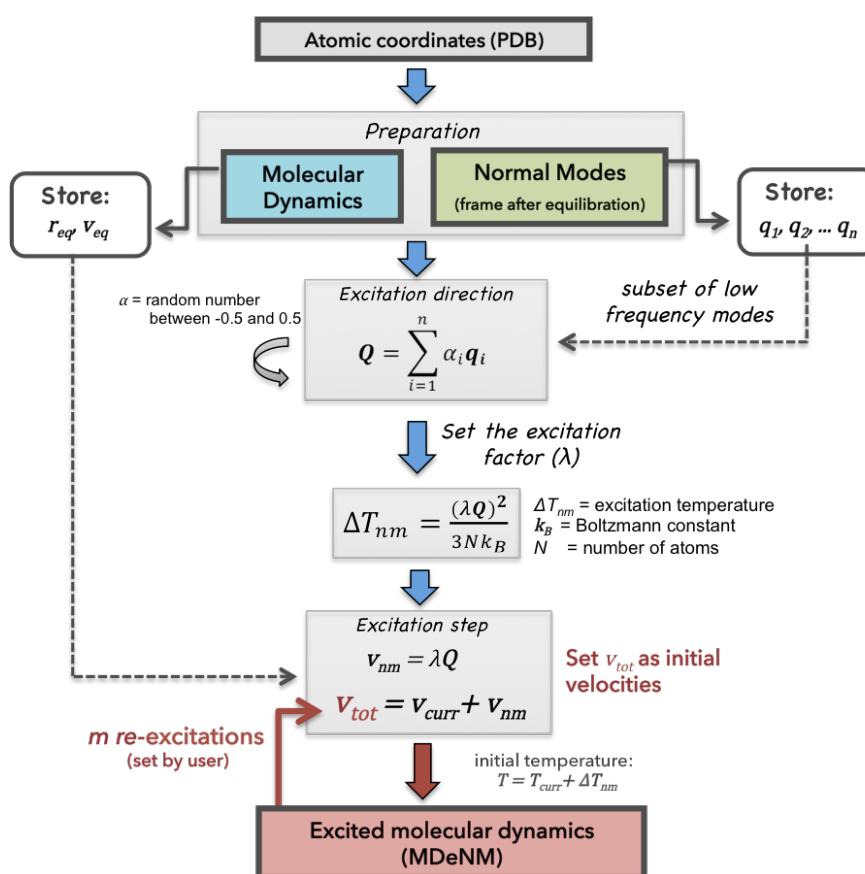


Figure 1: Flowchart of the MDeNM procedure. After a preparation step (described in the present section), a random linear combination of NM vectors is achieved yielding the vector \mathbf{Q} , followed by the calculation of an excitation factor (λ) in relation to the additional temperature of the normal mode space (ΔT_{nm}) set by the user. Subsequently, the excitation velocity is assigned along \mathbf{Q} and added to the current velocities (v_{curr}). Finally, an excited dynamics simulation (MDeNM) using the new velocities is carried out. Prior to the first excitation, velocities and coordinates are obtained from equilibration MD, whereas upon re-excitations they are obtained from the last step of the previous MDeNM simulation.

2.1.1. Preparation step

This step consists of two independent procedures: *i.* an equilibration MD simulation and *ii.* a NM calculation of the last frame. These calculations may be performed with varying protocols depending on the system simulated. Details of the procedures performed in this study are given in the Methods section.

Herein, after the equilibration MD, we store the final atomic velocities and positions obtained (\mathbf{v}_{curr} , \mathbf{r}_{eq}). Subsequently, we perform an energy minimization on the

equilibrated structure, and then normal modes are calculated. Vectors from the low frequency end of the vibrational spectrum are also stored ($\{q_i\}$).

2.1.2. Determination of the excitation direction

The first step in each MDeNM simulation is the determination of the direction to be excited. The direction \mathbf{Q} is defined as a linear combination of n NM vectors (q_i) as given by:

$$\mathbf{Q} = \sum_{i=1}^n \alpha_i q_i \quad [1]$$

α_i values correspond to uniformly distributed random numbers between -0.5 and 0.5. This procedure yields distinct directions to be excited in each replica simulation (Supplementary Figure 1). The resulting vector is thereafter normalized. It is important to stress here that the number of modes chosen does not affect the computational cost of each replica since always a single direction is subjected to the excitation procedure, but the number of replicas to be considered for a full coverage of conformations increases with the dimensionality of the NM space.

2.1.3. Excitation step

The extra temperature ΔT_{nm} related to the NM space must be sufficiently elevated to ensure the occurrence of collective motions described by normal mode vectors. Indeed, ΔT_{nm} is crucial in the MDeNM approach since it controls how fast and to what an extent the conformational space will be explored. Considering that a single degree of freedom (described by \mathbf{Q}) is excited in each replica, high values of ΔT_{nm} (e.g., above 50K) will result in increased acceleration of large amplitude motions, without significant relaxation due to other degrees of freedom. This important aspect will be further discussed in the

Results section. Aiming to reach different points in the NM space, the range of excitation temperatures is randomized:

$$\Delta T_{nm} = \Delta T_{user} \cdot \gamma \quad [2]$$

Here, ΔT_{user} is the maximum excitation temperature entered as an input parameter by the user, γ is a uniformly distributed random number between 0 and 1. This step is carried out before the excitation. Consequently, each replica simulation presents distinct values of ΔT_{nm} . This defines a range of excitation temperatures for the multi-replicas, therefore increasing the randomness of sampling along excited directions. Subsequently, an excitation factor λ is defined such that an assigned velocity along \mathbf{Q} yields the temperature ΔT_{nm} over all degrees of freedom:

$$\Delta T_{nm} = \frac{(\lambda \mathbf{Q})^2}{3Nk_B} \quad [3]$$

Where k_B is the Boltzmann constant and N the number of atoms of the system. The atomic masses are not explicitly introduced in the eq. 3 since normal modes, and therefore \mathbf{Q} vectors, are already weighted by the square root of atomic masses according to NM theory. Finally, the Cartesian atomic velocities related to the NM space ($\mathbf{v}_{nm} = \lambda \mathbf{Q}$) are added to the velocities of the last step of the previous simulation (\mathbf{v}_{curr}):

$$\mathbf{v}_{tot} = \mathbf{v}_{curr} + \mathbf{v}_{nm} \quad [4]$$

Next, multiple short MD simulations of few ps are restarted from the final frame of the equilibration using \mathbf{v}_{tot} as the initial velocities (varying the direction of \mathbf{Q} in each one).

2.2. Dissipation time of the excitation energy

To monitor the excitation effects during MDeNM simulations, we calculated the projections of the coordinates and mass-weighted velocities $\mathbf{v}_Q^m(t)$ onto \mathbf{Q} (here mode 7

of lysozyme as further described in the Results section). The corresponding autocorrelation functions (ACF) are also given in Figure 2. The mass-weighted velocity $v_Q^m(t)$ along \mathbf{Q} is defined as:

$$v_Q^m(t) = M^{1/2} \mathbf{v}(t) \cdot \mathbf{Q} \quad [5]$$

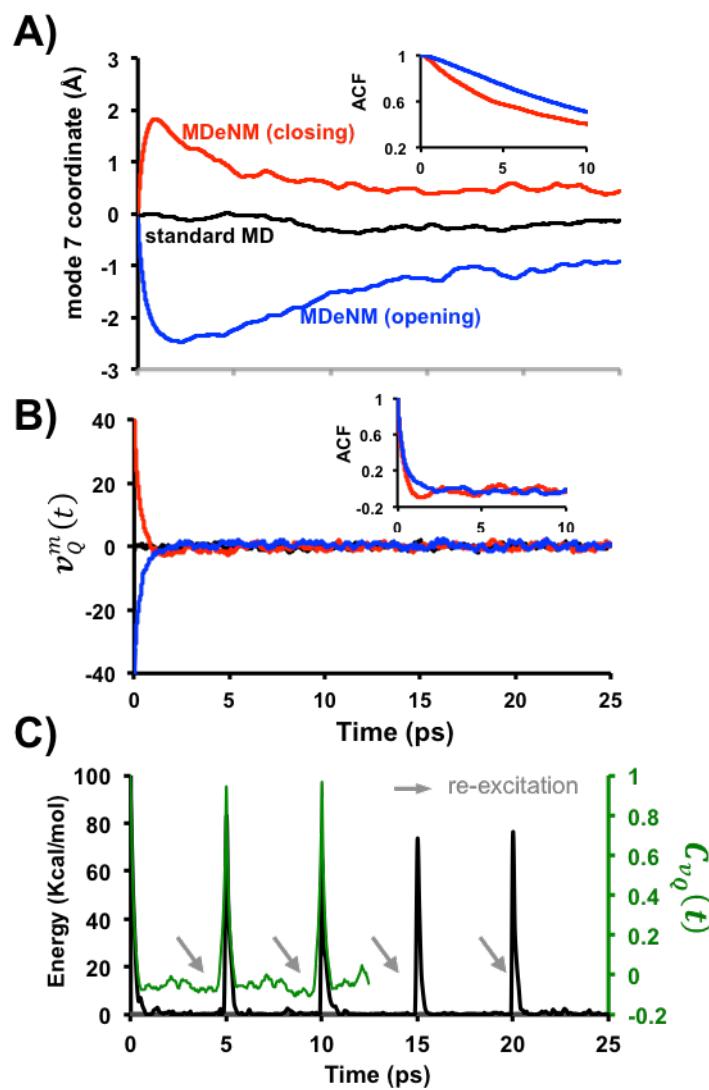


Figure 2: Monitoring the excitation effects during MDeNM simulations. Independent 25ps simulations were carried out in explicit solvent exciting ($\Delta T_{nm}=10K$) each direction of mode 7 of lysozyme (blue and red lines correspond to opening and closing motions respectively). **(A)** Positional relaxation of the NM coordinate. **(B)** Time evolution of the mass weighted velocities along the each direction of mode 7. The corresponding autocorrelation functions are given in the inset of each plot. The black lines represent results obtained with standard MD. Distinct behaviors

1
2
3 are observed (frequency of oscillations and decay lengths) depending on the excitation directions.
4
5 (C) The MDeNM multi-excitation scheme. Time evolution of the energy along the excited
6 direction calculated from 25 ps simulations carried out in explicit solvation (left axis). The
7 autocorrelation function $C_{vQ}(t)$ considering half of the simulation length is represented in green
8 (right axis). Re-excitations were performed each 5 ps as indicated on the graph.
9

10 As expected, velocity relaxation is faster than positional relaxation. (Figures 2A and B).
11
12 Even so, both take place on the picosecond timescale. These effects must be taken into
13 account when simulations are compared to experimental relaxation measurements.
14
15 However, the fast dissipation time of the excitation energies (Figure 2B) indicates that
16 high-energy conformations may be explored under periodic excitations without
17 perturbing the stability of the simulation.
18
19

20 We also noticed that excitation effects throughout MDeNM simulations are sensitive to
21 the kind of the enhanced motion. In this example, a displacement along the positive
22 direction (closing) relaxed faster than along the opposite direction related to opening.
23 This behavior may be related to (1) distinct energetic costs associated to these
24 displacements and (2) to a more efficient energy dissipation during closing because of
25 atomic collisions²¹.
26
27

28 To understand the sensitivity of $C_{vQ}(t)$ to simulation parameters (e.g. solvation
29 conditions such as vacuum, continuum model or explicit water), we compared the curves
30 calculated after MDeNM simulations carried out using distinct solvation models (see
31 Figure 2 in Supporting Information). As expected, $C_{vQ}(t)$ quickly decays in the explicit
32 solvent simulation due to the friction imposed by water molecules. In contrast, we
33 observed a slower decay with the Generalized Born continuum model²², and intermediate
34 values for calculations in vacuum.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2.2.1. Multi-excitation scheme

Enhancing large conformational changes with MDeNM might require a sufficiently high amount of energy injected along \mathbf{Q} , because of the fast decay time of excitation (especially in explicit water). However, this may also lead to unrealistic structural distortions. Alternatively, a multi-excitation scheme presents the advantage of exploring the conformational space more adiabatically, since low excitation energies can be used (Figure 2B). In order to avoid an excessive heating, two successive excitation steps should be performed in an interval larger than the decay time of $C_{v_Q}(t)$ in the simulation condition. The multi excitation approach is slower when compared to a single excitation procedure using a higher ΔT_{nm} . However, the number of excitation steps can be slightly increased to heat up the system if one desires to reach structures with higher energy levels ('excited structures') and/or to characterize metastable or transition states.

2.3. Conformational free energy exploration

Accurate estimates of conformational free energies can be obtained with a simple protocol as depicted in Figure 3. Initially a set of MDeNM replica simulations is performed and the resulting trajectories are merged into a single one. Then a clustering procedure is applied yielding a smaller subset of relevant conformations that will be submitted to standard MD simulations. In order to avoid remaining excitation effects it is preferable to run for each of these simulations a short equilibration run allowing the system to de-excite by itself. This might be important to generate unbiased trajectories starting from relevant conformations. Finally, free energy estimates can be obtained along any desired reaction coordinate in one dimension, or within a plane in multidimensional case (one or more modes for example) based on population histograms.

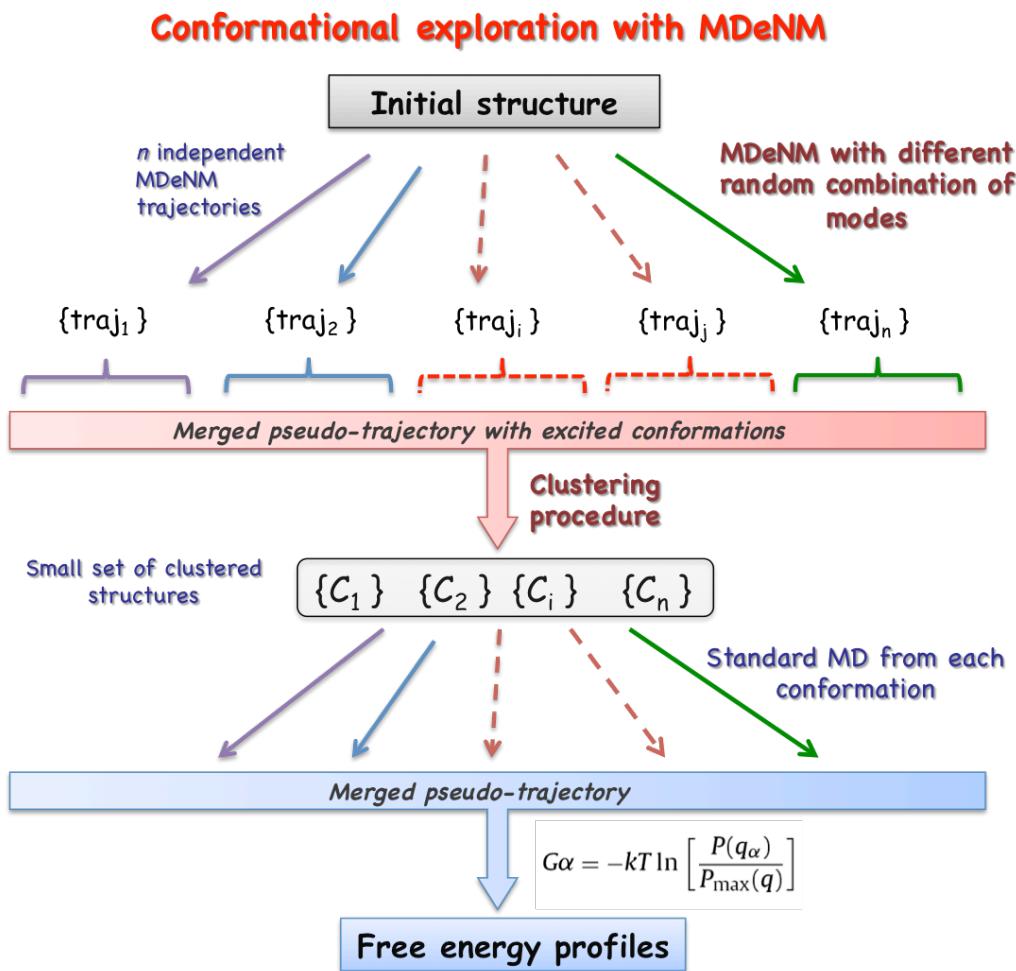


Figure 3: Schematic representation of the protocol to calculate conformational free energies. The initial steps of this protocol are described in the Figure 1 scheme.

3. Results

3.1. The hen egg white lysozyme – test case 1

Lysozyme (E.C 3.2.1.17) is one of the most studied proteins from a structural perspective²³. It was the first enzyme to have its structure solved²⁴ and since then, many structural studies reported important functional mechanisms²⁵⁻²⁷. Hen egg-white lysozyme (HEWL) is a highly studied form, being composed of 129 residues arranged into two domains (α and β). The α -domain contains five α -helices and the β -domain three

1
2
3 anti-parallel β -strands. Four disulphide bonds (three intra-domain and one inter-domain)
4
5 are crucial in stabilizing the tertiary structure (Supplementary Figure 3).
6
7
8

9 **3.1.1. Lysozyme collective motions described by normal modes**
10

11 Early NM studies provided a detailed picture of the hinge bending motion described by a
12 single vibrational mode ($\nu=3.68\text{cm}^{-1}$) that is related to substrate entrance²⁸. In addition,
13 the second lowest frequency mode was related to a twisting motion between the domains
14 ($\nu=4.78\text{cm}^{-1}$). Recent MD data indicated that such motions accounts for more than 70%
15 of HEWL overall dynamics, thus confirming their relevance²⁷.
16
17

18 Our results were in agreement with these studies since the two lowest frequency internal
19 modes (7 and 8) computed here corresponded to hinge bending ($\nu=3.64\text{ cm}^{-1}$) and
20 twisting ($\nu=4.62\text{ cm}^{-1}$) motions (Figure 4). The C_α squared fluctuations contributed by the
21 100 lowest frequency NMs were also similar to previously published data²⁸. Focusing on
22 fluctuations contributed by the two slowest modes (Figure 4C), we noted a striking
23 similarity with the profile obtained from the first 100 lowest frequency NMs (correlation
24 coefficient, $r=0.85$).
25
26
27
28
29
30
31
32
33
34
35
36
37
38

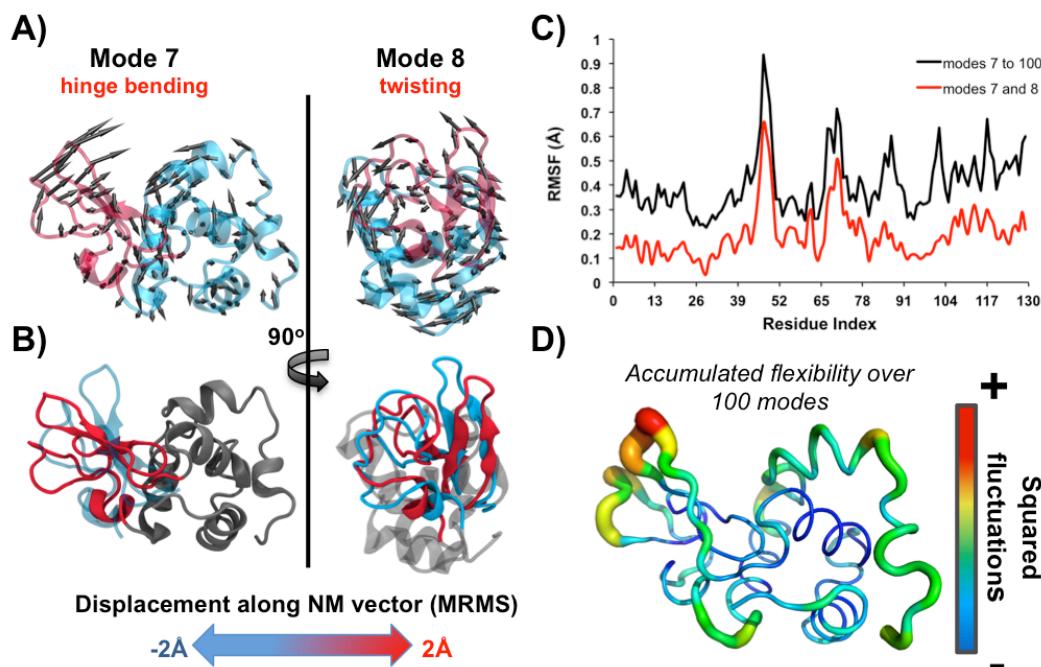


Figure 4: Collective motions described by low frequency normal modes of HEWL. **(A)** The directions of motions are represented by arrows and the magnitude is proportional to their amplitudes. The α and β -domains are colored blue and red, respectively. **(B)** The 2 Å MRMS displaced structures along both NM directions (- blue and + red), corresponding to relative displacements between the α and β -domains (hinge bending – mode 7, and twisting – mode 8). **(C)** Comparison between the RMSFs obtained with distinct sets of modes showing that modes 7 and 8 govern the overall pattern of fluctuations ($r = 0.85$). **(D)** Total flexibility contributed by the first 100 lowest frequency NMs embedded onto the lysozyme structure. The width of the tubes and their colors are proportional to the magnitude of the total square fluctuations.

3.1.2. Structural analysis of MDeNM simulations along one mode

After calculation of HEWL normal modes we performed independent MDeNM and standard MD simulations starting from the same initial conformation. Here, the hinge bending motion described by mode 7 was excited (either in the closing or opening directions). Figure 5 displays the time evolution of the excitation energy injected along mode 7 given by $\frac{1}{2} \mathbf{v}_Q^m \cdot \mathbf{v}_Q^m$, and also the associated RMSD, radius of gyration and hydrophobic surface variations. Clearly, large structural modifications occur in the first 5 ps after excitation. Although the excitation energy rapidly decreases to values close to zero, this single pulse of excitation results in an increase of RMSD (up to 2-4 gh)

followed by a subsequent stabilization, contrasting with the low values corresponding to the standard MD trajectory (black lines). This finding suggests the system is trapped within minimum energy wells after excitation. The variation of the radius of gyration was directly related to the hinge bending motion described by mode 7; it increased during opening and decreased during closing.

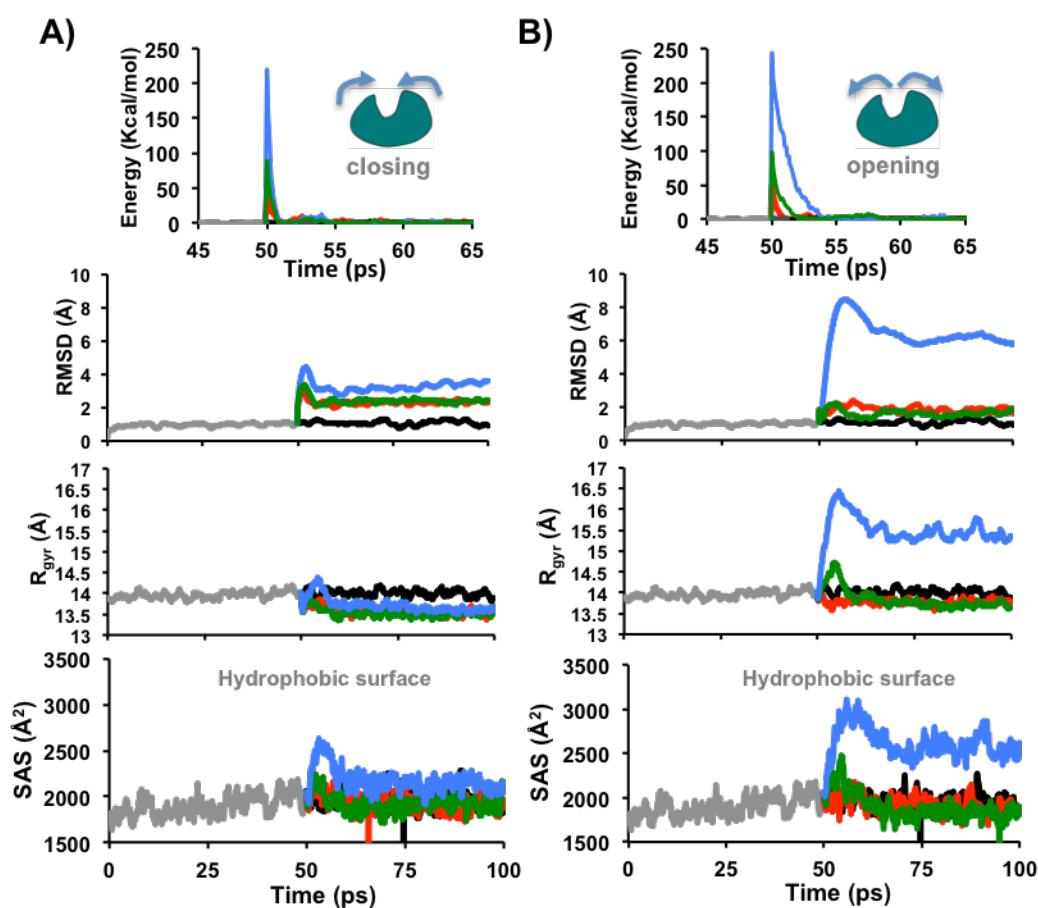


Figure 5: Kinetic energy along the mode, RMSD, radius of gyration and hydrophobic surface variations computed from standard MD and MDeNM simulations carried out for HEWL in vacuum with a distance dependent dielectric constant. The excitation was performed at t = 50 ps. (A) Closing motion upon excitation of the positive direction of mode 7 in independent MDeNM continuation runs using distinct ΔT_{nm} values (red: 10K, green: 20K, blue: 50K). The black curves

1
2
3 displays the same properties obtained from a standard MD starting from the same conformation.
4
5 (B) Same as in A but exciting in the reversed direction of mode 7 (opening).
6
7

8 Notably, exciting along the positive direction (closing) at $\Delta T_{nm} = 50$ K resulted in an
9 increase of hydrophobic surface exposure at 55 ps (Figure 5A), suggesting an important
10 structural perturbation, subsequently stabilized at a longer time (~100 ps). Accordingly,
11 the RMSD between snapshots ($t=100$ ps) obtained after MDeNM trajectories with $\Delta T_{nm} =$
12 20 or 50 K did not reveal striking differences (RMSD = 1.27 Å) showing convergence of
13 both simulations to produce similar structures. On the other hand, a huge increase in the
14 RMSD, Rg values and exposure of hydrophobic surface is observed upon excitation of
15 the opening motion at $\Delta T_{nm} = 50$ K (Figure 5B). These observations may be related to an
16 irreversible unfolding mechanism with crossing of an energy barrier only when a high
17 input of kinetic energy (50 K) was injected along mode 7. Overall, these results show that
18 special attention has to be brought about the excitation level in order to preserve the
19 global protein structure.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 3.1.3. Extents of sampling reveals intrinsic lysozyme dynamics 38

39 Standard NM analysis describes relatively fast vibrational motions along normal mode
40 vectors with corresponding amplitudes substantially lower than those experimentally
41 observed²⁹. One of the advantages of the MDeNM approach is the possibility to excite a
42 large set of modes simultaneously, allowing the determination of the extents of sampling
43 based only on intrinsic dynamical features. To test the method, we carried out 200 x 2 ps
44 MDeNM replicas exciting the first modes of the vibrational spectrum (modes 7 to 20),
45 yielding an accumulated simulation time of 0.4 ns. For comparison purposes, we also
46 performed two independent 100 ns standard MD simulations. These simulations were
47
48
49
50
51
52
53
54
55
56
57
58
59
60

carried out at constant temperature (300 K) using Generalized Born (GB) implicit solvation model. Overall, we observed a greater extent of sampling by mode obtained from MDeNM trajectories when compared to standard MD (Figure 6 and Supplementary Figure 4). Regarding the lowest-frequency modes (7 to 10), differences are markedly larger. Although energy was not preferentially injected onto any particular direction, since only random linear combinations of modes were considered, the maximum MRMS deviation along a given mode underlies distinctive features of the energy surface related to each mode (how large it is). The broader distributions observed for modes 7 and 8 corroborate recent data indicating that such motions correspond to more than 70% of HEWL dynamics²⁷.

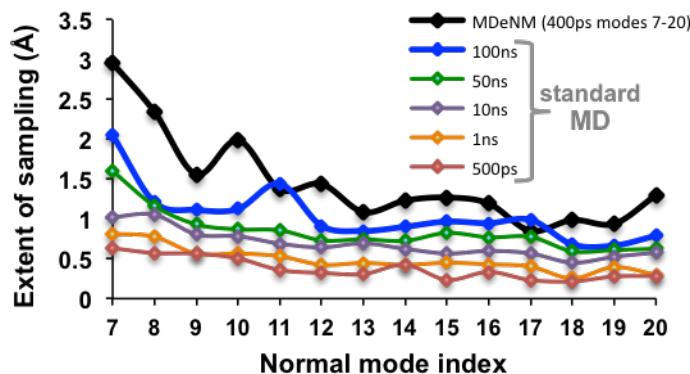


Figure 6: Displacement range, from the minimum to the maximum values (in MRMS scale) of NM coordinates for each low frequency mode (from 7 to 20) obtained from MDeNM and MD trajectories of HEWL. All NM coordinates were computed over each snapshot sampled during the simulations with respect to the reference structure used in NM calculations. Simulation lengths are indicated in the legend. (See Methods section for the calculation of the NM coordinates).

3.1.4. Comparing theoretical and experimental flexibilities

We initially compared in Figure 7A the atomic fluctuations obtained from MDeNM and MD simulations (same trajectories used to produce Figure 6). Although similar flexibility profiles were obtained from both computational methods, the magnitudes of MDeNM

1
2
3 fluctuations were markedly higher. It must be stated that 250 times more calculations
4 were employed to obtain such results using standard MD (0.4 ns MDeNM x 100 ns MD).
5
6

7 Further, the profiles obtained from both computational methods were similar to the
8 structural variation displayed by an ensemble of 501 HEWL crystal structures. They
9 presented equivalent Pearson correlation coefficients when compared to the experimental
10 data ($r_{MDeNM-Xray}=0.71$, $r_{MD-Xray}=0.69$). It must be considered that different crystal
11 environments affect the fluctuations obtained from the experimental data³⁰. However, the
12 dominant aspects of HEWL flexibility profile should be accessed when considering a
13 large number of structures in the ensemble, which enables a rough comparison with
14 simulation data.
15
16

17 Next, we investigated whether the RMSF obtained with MDeNM would be comparable
18 to the structural variation within a refined NMR ensemble using residual dipolar coupling
19 (RDC) data³¹. We observed similar trends, as confirmed by the correlation coefficient
20 obtained ($r_{MDeNM-NMR}=0.62$). Interestingly, the structural variations exhibited by both
21 experimental ensembles presented a higher correlation coefficient ($r_{X-ray-NMR}=0.78$).
22
23

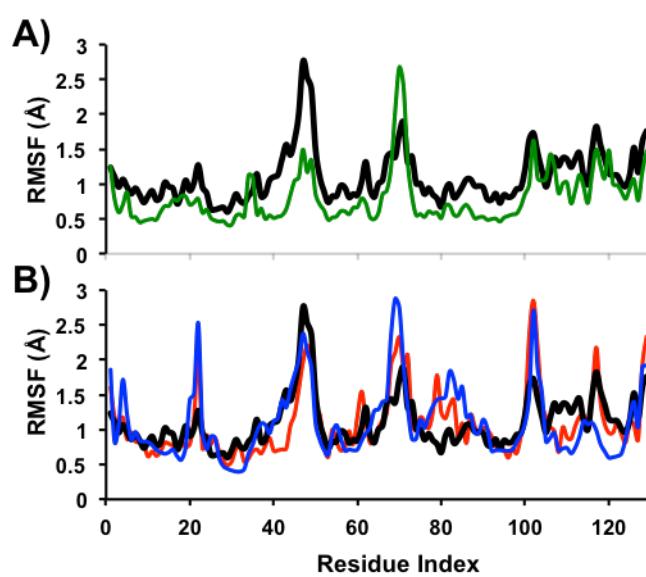


Figure 7: Comparing HEWL fluctuations obtained from distinct sources. **(A)** RMSF computed from simulations: 200 x 2 ps MDeNM simulations (0.4 ns total) upon excitation of modes 7 to 20 (black) and 100 ns standard MD simulations (green); **(B)** structural variation calculated from experimental ensembles: 501 X-ray structures and from a NMR RDC-refined ensemble (50 structures)³¹ are represented in blue and red respectively. The black line corresponds to the MDeNM simulation results given in *A*. To facilitate comparisons, the RMSF obtained each experimental ensemble was scaled to have the same mean as those obtained with MDeNM simulations.

We also investigated if motions described by both approaches were also compared to those obtained from 52 HEWL structures obtained by NMR³¹⁻³². From a PCA analysis on these distinct ensembles, the first five eigenvectors corresponding to higher eigenvalues were selected to define an essential subspace comprising at least 75% of the overall fluctuations. Table 1 presents the projections between normalized vectors extracted from NMR experiments and those obtained from conformations sampled by MDeNM and by MD simulations.

The examination of Tables 1 and 2 reveals that large amplitude motions extracted from MDeNM trajectories are not only more similar to those obtained from multiple NMR structures but also more collective than movements obtained from longer standard MD simulations. The PC1 from the NMR ensemble corresponds to the PC2 from MDeNM (twisting); while the NMR PC3 corresponds to the PC1 from MDeNM (hinge bending). The most collective PC from both NMR and MDeNM ensembles corresponds to the twisting motion. The second most collective motion is the hinge bending motion. However, it is important to consider that the structural variability present in the NMR ensemble reflects both thermal fluctuations and the lack of completeness of the model. Altogether, our comparisons between experimental data and MDeNM simulations, at

least allowed us to conclude that the excitation of multiple modes results in a better reproduction of experimental data than longer standard MD, which certainly is of interest when large systems are studied.

Table 1: Projections between vectors obtained from PCA of NMR structures, and those corresponding to a standard 200 ns MD and 0.4 ns MDeNM simulations in which modes 7 to 20 were excited. Absolute values of the projections are presented.

		Overlap between principal components				
		standard MD 200ns				
NMR ensemble	PC#	1	2	3	4	5
	1	0.03	0.37	0.11	0.31	0.27
	2	0.27	0.13	0.18	0.21	0.26
	3	0.02	0.37	0.04	0.14	0.07
	4	0.34	0.12	0.12	0.06	0.15
	5	0.18	0.27	0	0.07	0.02

		MDeNM (modes 7 to 20)				
NMR ensemble	PC#	1	2	3	4	5
	1	0.27	0.48	0.16	0.03	0.1
	2	0.12	0.17	0.49	0.08	0.03
	3	0.57	0.11	0.33	0.11	0.56
	4	0.03	0.02	0.03	0.07	0.09
	5	0.28	0.09	0.04	0.12	0.05

*highest values by PC highlighted in gray

Table 2: Degree of collectivity of motions described by the first PCs

PC#	Experimental	0.4ns MDeNM	200ns standard MD
1	0.6	0.43	0.27
2	0.42	0.67	0.36
3	0.37	0.39	0.21
4	0.45	0.71	0.54
5	0.18	0.54	0.43

3.1.5. Free energy landscapes along low frequency modes

FELs were computed by considering the first two lowest frequency modes (7 and 8) excited separately or simultaneously, and compared to results obtained with other techniques. These two modes were chosen because together they account of overall HEWL dynamics (Figure 4). Figure 8 displays the results obtained after carrying out a conformational exploration protocol (Figure 3, and additional details described in Supporting Information) in explicit solvent where modes 7 or 8 were excited separately. For each mode considered, 200 MDeNM simulations (replicas) of 2 ps length each were carried out, followed by a clustering of the obtained snapshots using the GROMOS algorithm with a RMS cut-off of 0.8 Å (Supplementary Figure 5). Subsequently, we submitted a set of 20 representative conformations (centroids) to 1 ns standard MD. All FELs were calculated using the last 800 ps from these last simulations, thus taking into account only equilibrated conformations, discarding or decreasing any remaining excitation effects. We also present the results obtained with metadynamics using the same vectors as collective coordinates and similar force fields.

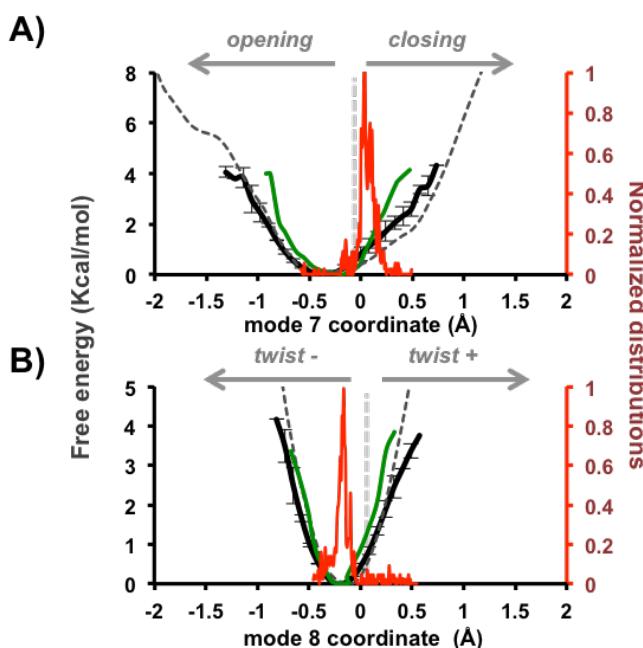


Figure 8: Free energy profiles along the two first normal modes of HEWL. Black lines correspond to the average (\pm SD) values calculated from two independent sets of standard MD started from representative conformations obtained with MDeNM exciting a single mode. Dashed curves correspond to results obtained with metadynamics. Green curves correspond to the FEL calculated from a 200 ns standard MD simulation. The normalized distribution of normal mode coordinates from 553 experimental structures is represented in red (right axis). A) Results obtained for mode 7 (hinge bending). B) Results obtained for mode 8 (+ and – corresponding to twisting towards the right and left hand sides, respectively).

The energy minimum was found at -0.3 \AA MRMS along mode 7, indicating a slight opened structure as the most favorable conformational state. Further, displacements from the minimum along both directions of mode 7 are allowed but with some dissymmetry. This observation is directly related to the distinct velocity relaxation curves observed in Figure 2A, where the decay of $C_{vQ}(t)$ was markedly slower for the opening direction. Notably, most of the coordinates extracted from 553 experimental structures lie on the positive side of the origin, therefore corresponding to closed states. This may be related to the presence of ligands and/or crystallization adjuvants, which perturb the natural equilibrium by favouring the stabilization of closed conformations. The small peaks

located at the minimum correspond to *apo* HEWL conformations. We obtained similar profiles with metadynamics, although it took 80 ns to converge, therefore slower than MDeNM, which required only ~20 ns. On the other hand, a longer 200 ns standard MD simulation was not sufficient to fully cover the hinge bending motion, indicating the need of much extended simulations in order to properly explore such slow movements.

Concerning mode 8, we obtained a narrower harmonic energy well (as expected due to its higher vibrational frequency in comparison to mode 7) and a smaller extent of sampling (-0.8 to 0.5 Å). Again, a 200 ns standard MD failed to properly account the whole extent of the FEL, although the results were closer to the MDeNM profile than for mode 7 (Figure 8B). Interestingly, most of the experimental structures lied within the energy minimum well. Similar results were obtained with metadynamics (45 ns to converge) showing that our method is equally accurate. In addition, our protocol required fewer calculations (~21 ns) to obtain the FELs for both modes studied here.

A bi-dimensional free energy landscape obtained with MDeNM simulations in which both modes 7 and 8 are simultaneously excited is shown in figure 9. It reveals similar extents of sampling along both modes as those obtained when a single mode was excited. However, the vast majority of X-ray and NMR structures were properly explored only with MDeNM: they are located within stable regions of the FEL up to 3 kcal/mol. The result obtained from 200 ns standard MD clearly shows that the simulation is trapped in a basin of the FES, unable to cross a barrier separating it from other relevant regions accessed in NMR experiments. In this respect, all HEWL structures obtained with NMR are in their *apo* form and therefore display a much larger structural variability (triangles) than those observed for the ligand bound X-ray structures (spheres).

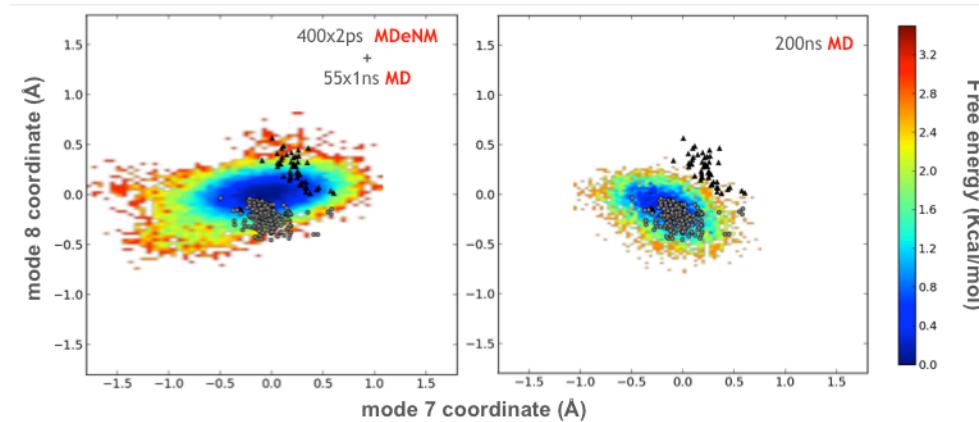


Figure 9: Bi-dimensional free energy profiles along the first two internal normal modes of HEWL. *Left:* FEL extracted from the distribution of conformations obtained from 55×1 ns standard MD started from activated conformations. *Right:* results obtained from a 200 ns classical MD. The gray circles and black triangles represent the normal mode coordinates for the X-ray and NMR ensembles, respectively.

3.2 The HIV-1 Protease - test case 2

The HIV-1 protease (PR) is another extensively studied protein that plays a crucial role on the virus life cycle³³. There are around five hundreds of experimentally determined available structures of this enzyme. This vast structural knowledge allows a survey of a huge number of conformations of PR complexes, with both inhibitors and substrates. PR is a symmetric homodimer containing 99 residues by chain arranged into three domains: core, terminal and flap (Figure 10A). While the former contains the catalytic residues, the second is composed by both N and C terminal segments that are possibly involved in dimerization. The flap domain is widely studied since it contain flexible hairpins covering the active site cleft and is involved on ligand binding by an open/close mechanism³⁴⁻³⁵.

3.2.1 HIV-1 protease flap dynamics described by consensus normal modes

Several experiments have shown that flap opening/closing dynamics occur on the $\mu s-ms$ timescales³⁴⁻³⁵, therefore impairing an appropriate study of this mechanism by standard MD simulations. NM analysis on a single PR structure also leads to a limited understanding of flap dynamics, as such motions may not be properly described. In this regard, our group developed a method to obtain a more accurate description of protein dynamics based on the calculation of normal modes from multiple minima, referred to as consensus modes (CM). Two of the low frequency CMs computed for PR (3 and 5) were observed to be directly related to flap opening³⁶. The vectors describing the CM3 and CM5 are represented in Figure 10B.

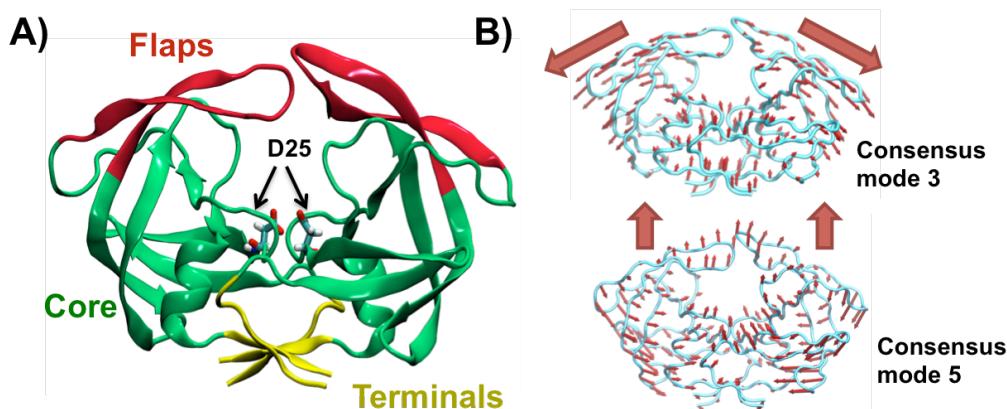


Figure 10: (A) Cartoon representation of PR structure (PDB ID: 1HHP) colored to represent its principal domains: terminal (yellow), core (green) and flaps (red). The catalytic residues D25 are highlighted in sticks. (B) Flap dynamics described by low frequency CMs. Displacements of C_{α} atoms along CM3 and CM5 are represented by arrows. The upper cartoon arrows represent the overall direction of motions of the flap domains. Adapted from ref.³⁶.

3.2.2 Quantitative correlation with experimental data

We investigate here whether the MDeNM approach allows a proper description of flap dynamics with fewer calculations. To this end, we conducted a similar conformational exploration protocol as for HEWL. Here, a clustering procedure was applied on the

conformations explored during 400×4 ps MDeNM simulations by exciting the CM3 ($\Delta T_{nm} = 10\text{K}$) related to the largest flap opening motion. The resulting 16 relevant structures (centroids obtained after clustering with a 0.8\AA cut-off) were further submitted to 1 ns standard MD simulations. All simulations were performed in explicit solvent. To obtain a quantitative comparison with experiment concerning flap opening, we compared the distribution of the distances between $\text{N}\zeta$ atoms of K55 from each monomer with that obtained by EPR experiments^{16,37}. This comparison is presented in Figure 11.

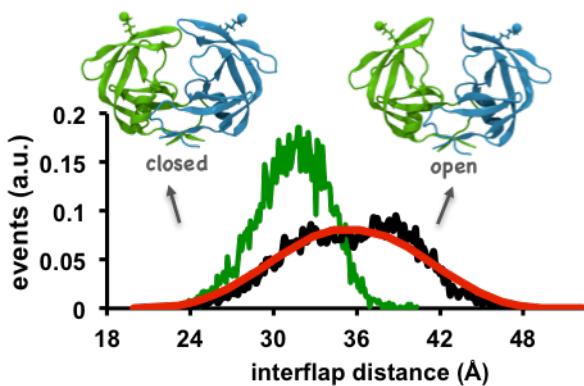


Figure 11: Quantitative analyses of flap dynamics. The distance between $\text{K55N}\zeta$ atoms from each monomer was used to measure the extent of opening/closing. Normalized distributions obtained with MDeNM conformational exploration protocol (as described in the text) and from EPR of a spin labeled PR are shown in black and red, respectively. The distribution obtained from a 200 ns standard MD simulation starting from a closed conformation is presented in green. Two extreme conformations obtained by 1.5\AA MRMS displacements along both directions of the CM3 are given in the upper part of the graph.

The distributions obtained from MDeNM simulations and from EPR present a very close similarity. In contrast, a 200 ns standard MD simulation was not sufficient to describe all of the conformational plasticity related to flap dynamics, since only closed and semi-open conformations were sampled. Together, these results reinforce the ability of MDeNM to provide an accurate description of protein slow dynamics. Similar results were previously obtained with umbrella sampling along the CM3 direction¹⁶.

3.2.3 Populations of accessible conformations

Figure 12 displays the bi-dimensional FEL obtained by carrying out 400×4 ps MDeNM simulations in which CM3 and CM5 were simultaneously excited, followed by clustering of the excited structures with a RMS cut-off value of 0.8 \AA , resulting in 40 relevant structures that were further submitted to 500 ps standard MD simulations. The extents of sampling obtained with MDeNM were larger along both modes than those obtained with standard 200 ns MD. Further, most of the 450 experimental conformations lied within 3.2 kcal/mol contour in the FEL obtained with MDeNM but not with MD. As previously discussed, the presence of ligands stabilizes conformations that are around 2-3 kcal/mol above the minimum, therefore characterizing a population shift toward closed structures (negative values along CM3 and CM5). The majority of the *apo* structures are close to the minimum energy region. On the contrary, the FEL obtained after a 200 ns standard MD simulation reveals trapping within a single minimum during the entire simulation.

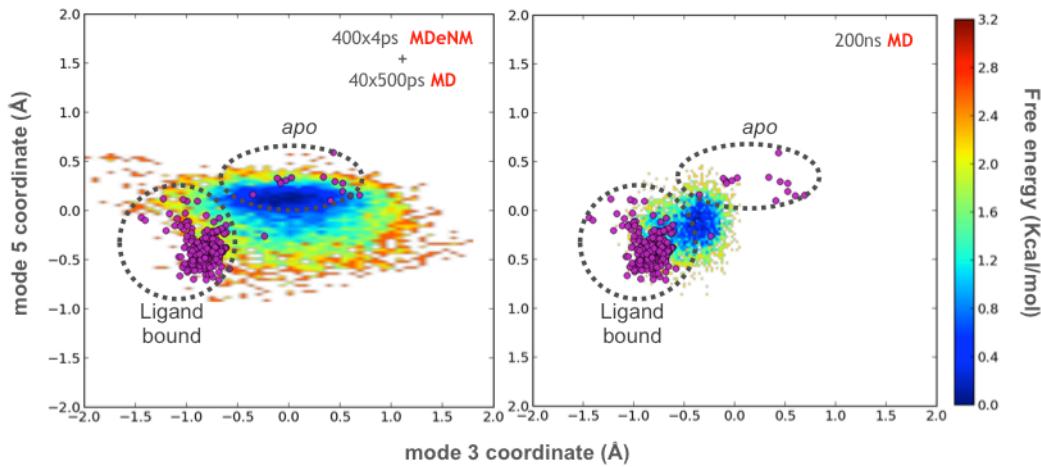


Figure 12. Bi-dimensional free energy profiles along both PR flap opening consensus modes. *Left:* FEL extracted from the distribution of conformations obtained from 40×500 ps standard MD from activated conformations. *Right:* results obtained from a 200 ns standard MD. The magenta circles represent the coordinates of 450 experimental structures.

4. Conclusions

Several hybrid simulation methods have been developed over the last years¹². Special attention has been devoted to approaches considering the intrinsic dynamics encoded in a given fold, especially normal mode analysis which appropriately describes such a feature³⁸⁻³⁹. Here, we presented an excited dynamics approach (MDeNM) to exhaustively explore the slow collective dynamics defined by the lowest frequency part of the vibrational spectrum and effectively deal with the combinatorial problem of a multidimensional space. Instead of only exploring single NM vectors as previously reported by others, multiple directions deriving from random linear combinations of low frequency NMs are kinetically excited in independent simulations. The large variety of energetically accessible conformations generated by our approach allows the determination of the extents of sampling in a complex collective coordinate space without arbitrary assumptions. Besides the statistical robustness provided by the multiple replica approach, MDeNM might be combined with clustering and standard MD simulations, thus allowing one to accurately compute free energy landscapes related to conformational changes.

The MDeNM is based in a multi-replica scheme with very short simulations (ps time scale), allowing effective parallelization. In terms of performance, our approach with only 0.4 ns (200 x 2 ps) for a two-dimensional normal mode space allows one to obtain a larger exploration of the ES compared to 200 ns of standard MD. It is at least three orders of magnitude more effective.

Finally, MDeNM results were closely correlated with experimental observations (X-ray, NMR, EPR) on both test systems considered, thus enabling its application to explore low-

1
2
3 populated excited states important to biological functions. Numerous applications can be
4 considered in the fields such as protein-ligand flexible docking, structural interpretation
5 of SAXS and spectroscopic data, and the conformational exploration of huge systems
6 (e.g. membrane proteins, viral capsids, molecular motors, etc.).
7
8
9
10
11
12
13

14 5. Methods 15

16 5.1 *Atomic coordinates and experimental datasets* 17

18 The following atomic coordinates were selected for all calculations performed in this
19 study: *apo*-lysozyme (PDB ID: 2LZT)⁴⁰ and *apo*-HIV-PR (PDB ID: 1HHP)⁴¹. A dataset
20 of 553 distinct hen-egg lysozyme experimental structures from the PDB (52 of them
21 being obtained by NMR and 501 by X-ray diffraction) was considered. The experimental
22 dataset of PR structures was the same as reported in³⁶. Only C_α atoms were considered
23 when calculating the projections of experimentally obtained structures onto NM vectors
24 for computing their normal mode coordinates.
25
26
27
28
29
30
31
32
33
34
35
36

37 5.2 *Force field* 38

39 All simulations were carried out using CHARMM v.36b1²⁰ or NAMD 2.9⁴² with the
40 CHARMM force field parameter set 22^{20,43}. Hydrogen atoms were added to the crystal
41 structures using the HBUILD routine of CHARMM. Van der Waals interactions were
42 calculated up to 10 Å, being approximated until 12 Å by using a switching function.
43 Electrostatic interactions were calculated up to 10 Å.
44
45
46
47
48
49
50

51 5.3 *Molecular Dynamics* 52

53 MD simulations were performed in distinct conditions of complexity: vacuum,
54 Generalized Born (GB) implicit solvent²² and explicit TIP3 water molecules⁴⁴. The first
55
56
57
58
59
60

1
2
3 tests were carried out in vacuum, in which a distance dependent dielectric constant
4
5 ($\epsilon=2r_{i,j}$) was employed to treat electrostatic interactions. Then, simulations were
6
7 performed representing the solute surrounded by a continuum electrostatics
8 representation using the GB algorithm (subroutine GBSW of CHARMM). Dielectric
9 constants for the protein interior and solvent were assigned values 1 and 80, respectively.
10
11 Lastly, simulations were performed in explicit solvent using periodic boundary
12 conditions and the PME algorithm to treat electrostatic interactions⁴⁵.
13
14

15 During the MD steps (heating, equilibration and production) SETTLE⁴⁶ and SHAKE⁴⁷
16 algorithms were used to fix the bonds within water molecules and protein allowing an
17 integration time step of 2 fs. Pressure and temperature were kept constant during
18 equilibration and production using the Berendsen algorithm⁴⁸⁻⁴⁹.
19
20

21 Briefly, systems were energy minimized using the conjugate-gradient algorithm keeping
22 protein heavy atoms harmonically restrained with a force constant of 50 kcal mol⁻¹ Å⁻² to
23 avoid artificial structural distortions. These restraints were progressively decreased to 2.5
24 during 500 ps of heating MD in which velocities were assigned accordingly to a
25 Maxwell-Boltzmann distribution corresponding to 50 K and then slowly increased to 300
26 K. Subsequently, during the equilibration phase, the positional restraints were
27 progressively decreased to zero during the first half of a 1ns constant temperature MD,
28 while in the remaining part all restraints were removed. Production simulations were
29 performed in the *NPT* ensemble.
30
31

50 51 52 53 54 55 56 57 58 59 60 5.4 Normal mode calculation

53 The consensus modes of the HIV-1 protease (PR) were calculated from 20 structures
54 from a 1 ns trajectory of the *apo* PR, starting with solvated crystal structure (PDB ID
55
56
57
58
59
60

1HHP), exactly as previously described in ref.³⁶. NM^s for lysozyme were calculated in
vacuum using a distance dependent dielectric constant ($\epsilon=2r_{ij}$), to treat electrostatic
interactions. Prior to NM calculations, the HEWL crystallographic structure was energy
minimized using the steepest descent (SD) and conjugate-gradient (CG) methods
followed by the Adopted Basis Newton Raphson (ABNR) algorithm. Harmonic restraints
were applied during the SD steps, that were progressively decreased from 250 to 0 kcal
 $\text{mol}^{-1}\text{\AA}^{-2}$. Subsequently, we further minimized the system with 1000 CG steps and then
applied the ABNR algorithm without positional restraints using a convergence criterion
of 10^{-5} kcal mol⁻¹Å⁻¹ RMS energy gradient. NM^s and atomic fluctuations were computed
with the VIBRAN module of CHARMM.

Upon NM calculation we used an analytical expression to compute the positional
fluctuations of each atom of the system²⁹. The contributions of n modes to the mean
square fluctuations of atom i , from the energy minimum position is:

$$\langle \Delta r_i^2 \rangle = k_B T \sum_{j=1}^n \sum_{\alpha=1}^3 \frac{q_{i\alpha,j}^2}{\omega_j^2} \quad [6]$$

Where k_B is the Boltzmann constant, T the absolute temperature, Δr_i the displacement of
atom i from its reference position in the energy minimized structure, $q_{i\alpha,j}$ is the element
corresponding to the i th atom in the j th normal mode vector, the index α ($\alpha=1,2,3$)
pointing to either of the three coordinate axes, and ω_j its frequency. The displacements
along the NM vectors q_j are expressed as a function of mass-weighted root mean square
(MRMS) distances as defined by the equation¹⁷:

$$d_{jMRMS} = \frac{1}{\sqrt{M}} \sum_{i=1}^{3N} \sqrt{m_i} (r_i - r_i^0) q_{ij} \quad [7]$$

where i indexes a given degree of freedom related to an atom with a mass m_i , M is the
total mass and q_{ij} is the i th element of the j th normal mode vector.

5.5 Degree of collectivity of a protein motion

The degree of collectivity of a protein motion⁵⁰ can be understood as a fraction of atoms participating to a given displacement. For a mode vector of length $3N$ with elements α_i , the degree of collectivity κ is defined as:

$$\kappa = \frac{1}{N} \exp\left(-\sum_{i=1}^{3N} \alpha_i^2 \log \alpha_i^2\right) \quad [8]$$

If the conformational change involves only a few atoms, κ is minimal. On the other hand, if $\kappa=1$ the given motion is maximally collective.

5.6 Metadynamics along NM coordinates

Metadynamics simulations were performed using NAMD 2.9⁴² collective variables (CV) module. Here only one-dimensional simulations were performed where a single mode was selected as CV during the 100 ns trajectories. To reconstruct the FEL related to displacements along NM vectors, we considered the portion of the simulation until the motion of the CV became diffusive. Therefore, each FEL obtained was related to a distinct simulation length because it depends on the nature of the motion explored. In particular, a Gaussian deposition rate of $0.01 \text{ kcal mol}^{-1} \text{ ps}^{-1}$ was used with a width of 0.1 \AA to explore the -2 to $+2 \text{ \AA}$ MRMS range. Two additional collective variables (using harmonic restraining potentials) were employed to prevent overall rigid rotation and translation of the protein thus maintaining the initial NM reference system

5.7 Free energy landscape (FEL) analysis

An *in-house* python script was developed to calculate the mono or bi-dimensional representation of the FEL. The free energy difference (ΔG_α) of a particular state α with

respect to the most populated one taken as reference, was calculated according to the probability of finding these states as given by:

$$G_\alpha = -k_B T \ln \left[\frac{P(q_\alpha)}{P_{\max}(q)} \right] \quad [9]$$

where k_B is the Boltzmann constant, T the temperature of the simulations, $P(q_\alpha)$ an estimate of the probability density function obtained from mono or bi-dimensional histograms of the projections of conformational changes sampled during MD onto NM vectors (q_i and q_j); $P_{\max}(q)$ is the probability of the most visited state. Bi-dimensional FEL were calculated from the joint probability distributions $P(q_i, q_j)$.

5.8 Autocorrelation Functions

Velocity autocorrelation functions were computed along the excited direction in MDeNM simulations. This analysis indicates the extent to which the system retains a “memory” of its previous values providing a useful measure of the relaxation times in MDeNM simulations. To facilitate the interpretation of data, the correlation function is normalized to a value of 1 at time zero and is expressed by:

$$C_{vQ}(t) = \frac{\langle (\mathbf{v}_m(t) \cdot \mathbf{Q})(\mathbf{v}_m(0) \cdot \mathbf{Q}) \rangle}{\langle (\mathbf{v}_m(0) \cdot \mathbf{Q})(\mathbf{v}_m(0) \cdot \mathbf{Q}) \rangle} \quad [10]$$

where \mathbf{v}_m and \mathbf{Q} are the mass weighted atomic velocities ($\mathbf{v}_m = \mathbf{M}^{1/2} \mathbf{v}$) and the excitation direction, respectively. Brackets indicate their time averages over the entire trajectory.

Associated Content

Supporting information. Figure S1: Uniform exploration of the conformational space described by NM coordinates during initial MDeNM steps. Figure S2: Velocity relaxation under distinct solvent representations. Figure S3: HEWL domain organization. Figure S4: Density plots showing populations of structures sampled during MDeNM simulations. Figure S5: Obtaining of representative structures with clustering. Additional details of the clustering procedure and an example MDeNM script are also given. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

Acknowledgments

All authors would like to thank the Brazilian-French collaborative project CAPES/COFECUB (project number: 678/10).

Author Information

Corresponding author: David Perahia, email address: david.perahia@ens-cachan.fr.

Note: The authors declare no competing interests.

6. References

1. James, L. C.; Tawfik, D. S., Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **2003**, *28*, 361-8.
2. Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E., Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Mol. Biol.* **2001**, *8*, 932-935.
3. Korzhnev, D. M.; Kay, L. E., Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **2008**, *41*, 442-51.
4. Berendsen, H. J.; Hayward, S., Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165-9.
5. Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H., Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. Rev. (Washington, DC, U. S.)* **2010**, *110*, 1463-97.
6. Grant, B. J.; Gorfe, A. A.; McCammon, J. A., Large conformational changes in proteins: signaling and other functions. *Curr. Opin. Struct. Biol.* **2010**, *20*, 142-7.
7. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.
8. Hess, B., Convergence of sampling in protein simulations. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2002**, *65*, 031910.
9. Hayward, S.; de Groot, B. L., Normal Modes and Essential Dynamics. In *Molecular Modeling of Proteins*, 2008; Vol. 443.
10. Skjaerven, L.; Hollup, S. M.; Reuter, N., Normal mode analysis for proteins. *Journal of Molecular Structure-Theochem* **2009**, *898*, 42-48.
11. Levy, R. M.; Perahia, D.; Karplus, M., Molecular dynamics of an alpha-helical polypeptide: Temperature dependence and deviation from harmonic behavior. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 1346-50.
12. Ho, B. K.; Perahia, D.; Buckle, A. M., Hybrid approaches to molecular simulation. *Curr. Opin. Struct. Biol.* **2012**, *22*, 386-393.
13. Schlick, T., Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *F1000Prime Reports* **2009**, *1*.
14. Kästner, J., Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 932-942.
15. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826-843.
16. Batista, P. R.; Pandey, G.; Pascutti, P. G.; Bisch, P. M.; Perahia, D.; Robert, C. H., Free Energy Profiles along Consensus Normal Modes Provide Insight into HIV-1 Protease Flap Opening. *J. Chem. Theory Comput.* **2011**, *7*, 2348-2352.
17. Floquet, N.; Durand, P.; Maigret, B.; Badet, B.; Badet-Denisot, M. A.; Perahia, D., Collective motions in glucosamine-6-phosphate synthase: influence of ligand binding and role in ammonia channelling and opening of the fructose-6-phosphate binding site. *J Mol Biol* **2009**, *385*, 653-64.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
18. Floquet, N.; Marechal, J. D.; Badet-Denisot, M. A.; Robert, C. H.; Dauchez, M.; Perahia, D., Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors. *FEBS Lett* **2006**, *580*, 5130-6.
 19. Bahar, I.; Atilgan, A.; Demirel, M.; 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.
 20. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009**, *30*, 1545-614.
 21. Perahia, D.; Levy, R. M.; Karplus, M., Motions of an Alpha-Helical Polypeptide - Comparison of Molecular and Harmonic Dynamics. *Biopolymers* **1990**, *29*, 645-677.
 22. Im, W.; Lee, M. S.; Brooks, C. L., 3rd, Generalized born model with a simple smoothing function. *J Comput Chem* **2003**, *24*, 1691-702.
 23. Johnson, L. N., The early history of lysozyme. *Nat Struct Biol* **1998**, *5*, 942-4.
 24. Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R., Structure of Hen Egg-White Lysozyme - a 3-Dimensional Fourier Synthesis at 2a Resolution. *Nature* **1965**, *206*, 757-&.
 25. Diehl, M.; Doster, W.; Petry, W.; Schober, H., Water-coupled low-frequency modes of myoglobin and lysozyme observed by inelastic neutron scattering. *Biophys J* **1997**, *73*, 2726-32.
 26. Yoshida, Y.; Ohkuri, T.; Kino, S.; Ueda, T.; Imoto, T., Elucidation of the relationship between enzyme activity and internal motion using a lysozyme stabilized by cavity-filling mutations. *Cell Mol Life Sci* **2005**, *62*, 1047-55.
 27. De Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M., Characterization of the interdomain motions in hen lysozyme using residual dipolar couplings as replica-averaged structural restraints in molecular dynamics simulations. *Biochemistry* **2013**, *52*, 6480-6.
 28. Brooks, B.; Karplus, M., Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc Natl Acad Sci U S A* **1985**, *82*, 4995-9.
 29. Brooks, B.; Karplus, M., Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 6571-5.
 30. Hinsen, K., Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics* **2008**, *24*, 521-528.
 31. Schwalbe, H.; Grimshaw, S. B.; Spencer, A.; Buck, M.; Boyd, J.; Dobson, C. M.; Redfield, C.; Smith, L. J., A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci* **2001**, *10*, 677-88.
 32. Refaee, M.; Tezuka, T.; Akasaka, K.; Williamson, M. P., Pressure-dependent changes in the solution structure of hen egg-white lysozyme. *J Mol Biol* **2003**, *327*, 857-65.

- 1
2
3 33. Kohl, N. E.; Emini, E. A.; Schleif, W. A.; Davis, L. J.; Heimbach, J. C.; Dixon, R.
4 A.; Scolnick, E. M.; Sigal, I. S., Active human immunodeficiency virus protease is
5 required for viral infectivity. *Proc Natl Acad Sci U S A* **1988**, *85*, 4686-90.
6
7 34. Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A., Flap
8 opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and
9 their implications for function. *Structure* **1999**, *7*, 1047-55.
10
11 35. Tozzini, V.; Trylska, J.; Chang, C. E.; McCammon, J. A., Flap opening dynamics
12 in HIV-1 protease explored with a coarse-grained model. *J Struct Biol* **2007**, *157*, 606-15.
13
14 36. Batista, P. R.; Robert, C. H.; Marechal, J. D.; Hamida-Rebai, M. B.; Pascutti, P.
15 G.; Bisch, P. M.; Perahia, D., Consensus modes, a robust description of protein collective
16 motions from multiple-minima normal mode analysis--application to the HIV-1 protease.
17 *Phys Chem Chem Phys* **2010**, *12*, 2850-9.
18
19 37. Galiano, L.; Bonora, M.; Fanucci, G. E., Interflap distances in HIV-1 protease
20 determined by pulsed EPR measurements. *J Am Chem Soc* **2007**, *129*, 11004-5.
21
22 38. He, J.; Zhang, Z.; Shi, Y.; Liu, H., Efficiently explore the energy landscape of
23 proteins in molecular dynamics simulations by amplifying collective motions. *The
Journal of Chemical Physics* **2003**, *119*, 4005-4017.
24
25 39. Gur, M.; Madura, Jeffry D.; Bahar, I., Global Transitions of Proteins Explored by
26 a Multiscale Hybrid Methodology: Application to Adenylate Kinase. *Biophys. J.* **2013**,
105, 1643-1652.
27
28 40. Ramanadham, M.; Sieker, L. C.; Jensen, L. H., Refinement of Triclinic Lysozyme
29 .2. The Method of Stereochemically Restricted Least-Squares. *Acta Crystallographica
Section B-Structural Science* **1990**, *46*, 63-69.
30
31 41. Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J., The 3-
32 Dimensional Structure of the Aspartyl Protease from the Hiv-1 Isolate Bru. *Biochimie*
33 **1991**, *73*, 1391-1396.
34
35 42. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.;
36 Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with
37 NAMD. *J Comput Chem* **2005**, *26*, 1781-802.
38
39 43. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.;
40 Karplus, M., CHARMM: a program for macromolecular energy,minimization, and
41 dynamics calculations. *Journal of Computational Chemistry* **1983**, *4*, 187-217.
42
43 44. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.,
Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.*
44 **1983**, *79*, 926-935.
45
46 45. Ulrich, E.; Lalith, P.; Max, L. B.; Tom, D.; Hsing, L.; Lee, G. P., A smooth
particle mesh Ewald method. *J Chem Phys* **1995**, *103*, 8577-8593.
47
48 46. Miyamoto, S.; Kollman, P. A., Settle - an Analytical Version of the Shake and
Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952-962.
49
50 47. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the
51 cartesian equations of motion of a system with constraints: molecular dynamics of n-
alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
52
53 48. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J.,
54 Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*,
55 Pullman, B., Ed. Springer Netherlands: 1981; Vol. 14, pp 331-342.
56
57
58
59
60

- 1
2
3 49. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak , J.
4 R., Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, *81*,
5 3684-3690.
6 50. Tama, F.; Sanejouand, Y. H., Conformational change of proteins arising from
7 normal mode calculations. *Protein Eng* **2001**, *14*, 1-6.
- 8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table of Contents (TOC) Graphics

