MDexciteR: Enhanced Sampling Molecular Dynamics by Excited Normal Modes or Principal Components Obtained from Experiments

Mauricio Costa, Paulo R Batista, Antoniel Gomes, Leonardo S Bastos, Maxime Louet, Nicolas Floquet, Paulo M Bisch, David Perahia

MDexciteR: Enhanced-Sampling Molecular Dynamics by Excited Normal Modes or Principal Components Obtained from Experiments

Mauricio G.S Costa[1, 2*], Paulo R. Batista[1], Antoniel Gomes[3], Leonardo S. Bastos[1], Maxime Louet[4], Nicolas Floquet[4], Paulo M. Bisch[3] & David Perahia[2*]

[1] Fundação Oswaldo Cruz, Programa de Computação Científica, Vice-Presidência de Educação, Informação e Comunicação. Av. Brasil 4365, Residência Oficial, Manguinhos. 21040-900, Rio de Janeiro / Brasil.

[2] École Normale Supérieure Paris-Saclay, UMR 8113, CNRS, Laboratoire de Biologie et de Pharmacologie Appliquée (LBPA), 4 Avenue des Sciences, 91190 Gif-sur-Yvette, France.

[3] Laboratório de Física Biológica, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

[4] IBMM, Univ Montpellier, CNRS, ENSCM, Montpellier, France.

E-mail addresses:

MGSC: mauricio.costa@fiocruz.br
PRB: paulo.batista@fiocruz.br
AG: antoniel.gomes@biof.ufrj.br
LSB: leonardo.bastos@fiocruz.br
ML: maxime.louet@umontpellier.fr
NF: nicolas.floquet@umontpellier.fr
PMB: pmbisch@biof.ufrj.br
DP: david.perahia@ens-paris-saclay.fr

*To whom correspondence should be addressed:

mauricio.costa@fiocruz.br & david.perahia@ens-paris-saclay.fr

Abstract

Molecular Dynamics with excited Normal Modes (MDeNM) is an enhanced sampling method for exploring conformational changes in proteins with minimal biases. The excitation corresponds to injecting kinetic energy along normal modes describing intrinsic collective motions. Herein we developed a new automated open-source implementation, MDexciteR, enabling the integration of MDeNM with two commonly used simulation programs with GPU support. Second, we generalized the method to include the excitation of principal components calculated from experimental ensembles. Finally, we evaluated whether the use of coarse-grained normal modes calculated with elastic network representations preserved the performance and accuracy of the method. The advantages and limitations of these new approaches are discussed based on results obtained for three different protein test cases: two globular and a protein/membrane system.

Introduction

The dynamic nature of biological macromolecules is widely evidenced by multiple sources of experimental data [1, 2]. Protein motions range from fast equilibrium fluctuations (picoseconds to nanoseconds) to slow conformational transitions of domains occurring in longer timescales (*i.e.*, microseconds, milliseconds). Multiple evidence supports the correlation between slow collective motions and biological functions [3, 4]. Furthermore, recent experimental data provided quantitative demonstrations of how such motions impact the conformational selection of states with optimal activity [5, 6].

Normal modes (NM) are intrinsic motions encoded by a given protein fold [3]. They often correlate with experimentally-determined structural transitions [7]. Several NM-based simulation techniques are currently available, some of them designed to elucidate the transition pathways between two known conformational states [8-13]. Other methods enable a more efficient exploration of the collective coordinate space without requiring endpoints. In this case, the subspace defined by selected NMs is extensively sampled through energy minimizations or molecular dynamics simulations [14-17].

In addition to NM analysis, Principal Component Analysis (PCA) is a commonly used technique to characterize conformational changes. The subspace spanned by a few degrees of freedom (first principal components - PCs) captures a large fraction of the overall variance of a given structural ensemble [18]. Because of the fast convergence of the essential subspace [19], PCA-based enhanced sampling methods allow an efficient conformational exploration during the time-course of MD simulations [20-26]

Our group has been developing NMA-based hybrid simulation approaches over the last decade [14, 17, 27]. Molecular Dynamics with Excited Normal Modes (MDeNM) stands out among then. This method allows an improved characterization of the conformational equilibrium with minimum computational costs. The broad range of MDeNM applications in biological systems include the determination of free energy profiles, structural rationalization of experimental data, flexible fitting of structures into Cryo-EM maps and ensemble docking studies [14, 28-31].

The motivation behind the MDeNM method is that standard MD simulations rarely explore the intrinsic motions described by low-frequency NMs. MDeNM was primarily designed to enhance the exploration of the conformational space defined by

a set of NMs without restrains or modifications in the energy function. This effect is achieved upon the kinetic excitation of collective motions during short MD simulations in the picosecond timescale. In other words, the directions of NMs are taken as a small increment to the current velocities of the simulated system. A multi-replica procedure is adopted to maximize MDeNM efficiency, where each independent simulation explores a distinct combination of selected NMs. The resulting amplitude of each excited movement is ultimately defined by the energy barriers found in the energy landscape.

In brief, the fundamental advantages of MDeNM are: *(i)* coupling between slow collective degrees of freedom and *(ii)* simulations performed on the same conditions used in standard MD, therefore enabling straightforward comparisons. According to previous evidence, the structural ensembles obtained by MDeNM encompass a much larger exploration of the conformational space than long standard MD [14, 28, 32].

Recently, our group participated in a critical assessment of NM-based hybrid methods [33]. All tested approaches exhibited increased computational efficiency without compromising their accuracy, based on comparisons with experimental and MD data. Furthermore, MDeNM performed well according to all metrics considered in the study. The authors suggested that the adoption of simplified elastic network models (ENM) for NM calculation instead of using all-atom energy functions would enhance MDeNM computational efficiency even further [33].

Based on these premises, we decided to generalize the strategy proposed for MDeNM by including the excitation of *(i)* NMs calculated using ENMs with distinct levels of coarseness; or *(ii)* PCs describing structural variations encoded in experimental ensembles. To accomplish this task, we developed a generalized automated implementation of MDeNM, termed MDexciteR. It is based on the integration of the previous MDeNM method into the R program (using the Bio3D package [34, 35]) in conjunction with a MD simulation engine (Figure 1). The advantages and limitations of each approach will be discussed in light of the results obtained for three different test cases.

2. Results

The enhanced sampling simulations reported in this study were conducted using our new implementation MDexciteR. Indeed, the *i/o* routines, NM or PCA calculations, and the linear algebra operations necessary to generate excitation velocities are performed within the R program [36]. In this article all simulations were carried out with the NAMD software [37], but it is important to notice that MDexciteR also enables integration with CHARMM [38]. Ongoing developments include AMBER [39] and GROMACS [40] compatibility. A detailed description of the implementation is given in the Supporting Information.

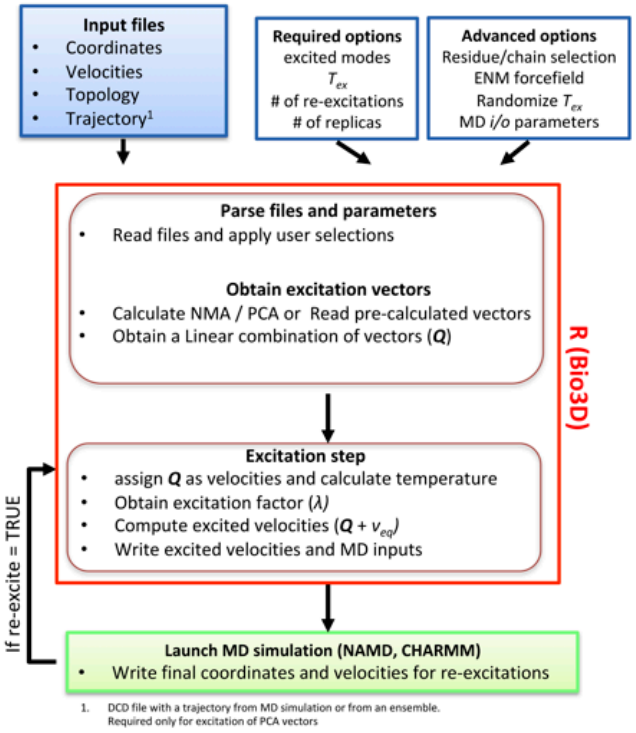

Figure 1: Workflow of the MDexciteR implementation. Blue boxes indicate user inputs; red, the routines conducted by the R software and in green, the execution of MD simulation, using previously generated excitation velocities.

We selected three proteins as test cases (Figure 2), including two soluble proteins and one membrane protein. First, the sampling efficiency of simulations guided by coarse-grained NMs or PC vectors was compared to the original method in simulations of lysozyme and calmodulin. Second, we tested the ability of MDeNM using C_α-based NMs to predict activation motions of a G-protein coupled receptor simulated in an explicit membrane environment.

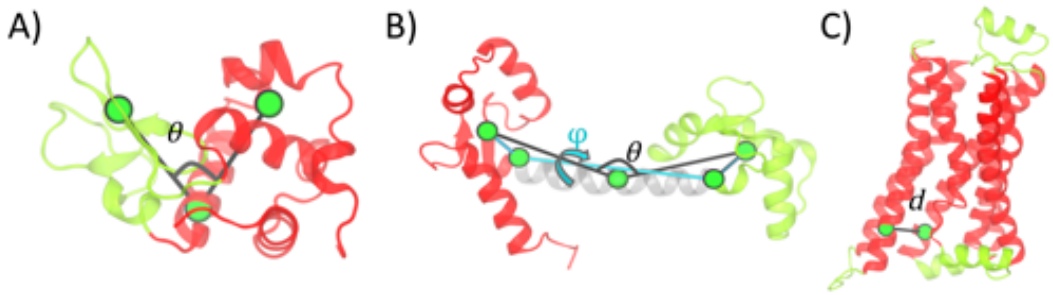

Figure 2: Test cases. (A) Hen Egg White Lysozyme (HEWL). The α and β domains are shown in *red* and *green*, respectively. The pincer angle θ is also highlighted. (B) Calmodulin (CAM). Each lobe is colored differently and the central helix is shown in *gray*. The bending (θ) and torsion (φ) angles are shown in *black* and *blue*, respectively. (C) *β2*-adrenergic receptor (*β2*AR) with the seven transmembrane helices shown in *red*. The distance between helix 3 and helix 6 (d) is also highlighted. All details about the structural descriptors used in this study are given in the methods section.

2.1 Test case 1: Lysozyme.

The high amount of available structural information makes Hen Egg White Lysozyme (HEWL; E.C 3.2.1.17) a strategic choice for testing new methodologies. In addition, lysozyme inter-domain motions are well described by the two lowest frequency normal modes (mode 7: hinge bending; mode 8: twisting) [41]. Here, we compare the extension of the conformational exploration using NMs calculated under distinct levels of complexity.

To enable fair comparisons, we initially obtained an equilibrated structure after a minimization/equilibration protocol in explicit solvent (as detailed in the methods section). Then, three independent NM calculations were carried out under decreasing levels of description: a "classical" all-atom approach with the CHARMM 36 forcefield [42] or using two distinct elastic network models (AAENM [43] and HCA [44]). While the AAENM model is an atomistic ENM, where nodes are placed in the positions of heavy atoms, the HCA model includes an additional level of simplicity, as it only considers C_α atoms as nodes. Both ENMs are based on distance-dependent force constants as previously described [43, 44].

Supplementary Figure 1 shows the overlap between pairs of normal modes calculated with each ENM and those obtained using the all-atom CHARMM 36 energy function. The degree of similarity to all-atom classical normal modes was

dependent on the level of coarseness: the atomistic AAENM model resulted in higher overlaps than the C_α-based HCA model. In general, the subspaces spanned by the ENM derived vectors were similar to those defined by all-atom NMs (as given by the root mean square inner product - RMSIP = 0.84 for AAENM and 0.82 for HCA), calculated according to the following equation:

$$RMSIP = \left(\frac{1}{5} \sum_{i=1}^{5} \sum_{j=1}^{5} (\boldsymbol{M_i} \cdot \boldsymbol{X_j})^2 \right)^{1/2}$$

where $\boldsymbol{M_i}$ and $\boldsymbol{X_j}$ correspond to NM and PC vectors obtained from distinct calculations. RMSIP values close to 1 indicate high similarity between the compared subspaces.

The comparison between NM vectors with Principal Components (PCs) calculated from an ensemble of 814 experimental structures was also performed (details in the Methods section). Figure 3 reveals low overlaps independent of the set of NM model considered. The only exception was the high overlap between PC5 and the lowest frequency NM 7 (markedly for CHARMM 36 and AAENM). Inspection of these vectors reveals a hinge bending motion in agreement with previous data [14] (Figure 3).

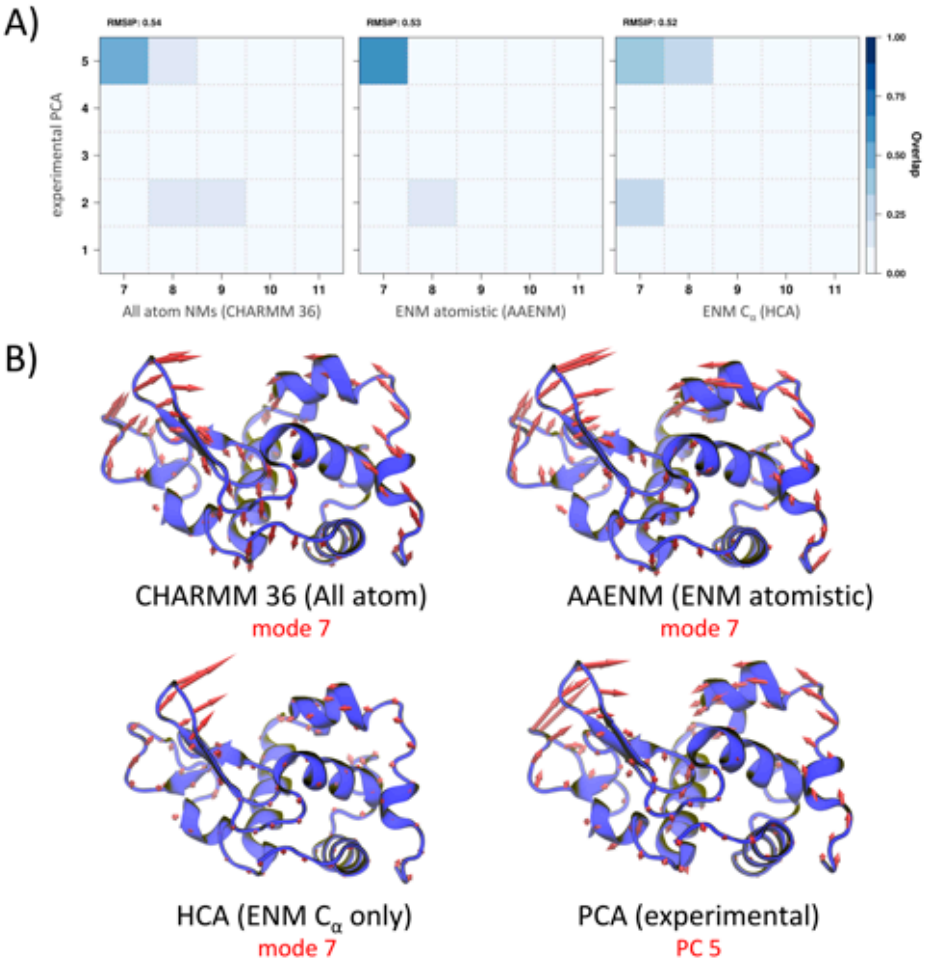

Figure 3: Comparing low-frequency NMs and principal components describing structural variations in the lysozyme experimental ensemble. (A) Overlap between PCs and different sets of NMs obtained with an all-atom model (CHARMM 36), an atomistic ENM (AAENM) or a C_α-ENM (HCA). (B) The directions of motions described by selected modes are represented by arrows, and the magnitude is proportional to their amplitude.

MDeNM simulations were carried out under an explicit solvent representation using the new MDexciteR implementation. The three sets of NM vectors obtained before using different models were taken as inputs. In addition, the top five PCs were also tested. One hundred replicas were carried out for each set of excitation directions considered (CHARMM 36, AAENM, HCA and PCA). Regardless of the NM model considered, we observed almost indistinguishable differences among the distributions of conformations. The range of exploration of all excited simulations was markedly greater than that obtained from a microsecond-long standard MD. Although the PCA-based exploration yielded a distribution similar to those obtained using NMs, the sampling extent was reduced yet substantially larger than standard MD (Figure 4A).

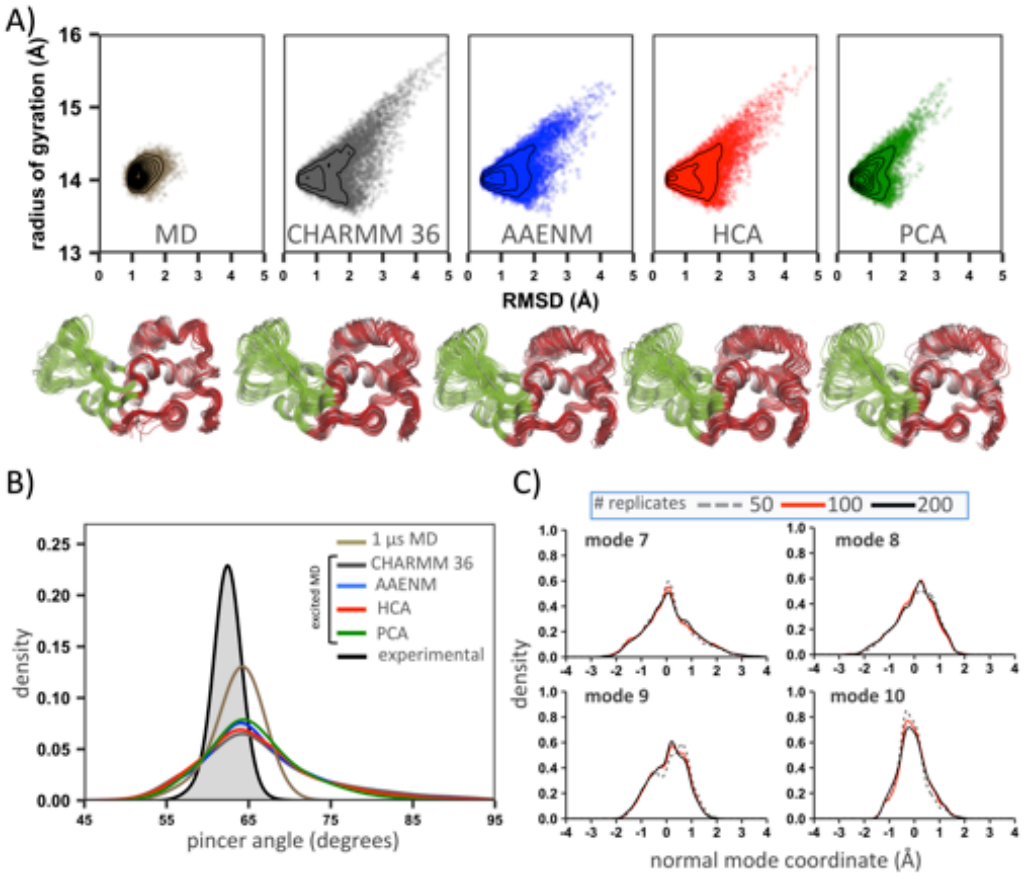

Figure 4: Enhanced conformational exploration using MDexciteR. (A) Bidimensional plots describing the conformational space explored as indicated in the legend. Both RMSD and radius of gyration were computed considering only C_α atoms, taking the starting structure as reference. The accumulated simulation times were 12 ns for each case. The distribution obtained after a 1 µs standard MD is given for comparison purposes. (B) Comparing the distribution of the pincer angle provided by an ensemble of 814 experimental conformers and simulations. (C) Projections of snapshots collected from simulations (MDeNM-HCA) along mass-weighted normal mode coordinates. Different numbers of replicates were considered to inspect convergence of the distributions.

The comparison between the conformational variability from simulations and the experimental ensemble revealed that MDeNM sampled the full range of observed pincer angle values (Figure 4B). The narrow distribution derived from the experimental conformers is likely related to crystal packing effects and the vast majority of ligand-bound states exhibiting restricted flexibility. We also checked the convergence of simulations with respect to the number of independent replicates (50, 100, and 200). One hundred replicates (selected for this study) seemed sufficient, as the resulting conformational populations were equivalent to those obtained with twice as many replicates (Figure 4C).

2.2 Test case 2: Calmodulin

Calmodulin (CAM) is a calcium-binding protein involved in diverse signaling pathways, acting as a Ca^{2+} dependent modulator of several enzymes, such as kinases or phosphatases [45]. CAM possesses two globular lobes that accommodate a pair of Ca^{2+} ions in each one. An extended flexible linker connects them, therefore yielding substantial structural plasticity. Indeed, CAM is a highly flexible protein found in a wide range of conformations, from extended to compact forms [46]. The intrinsic flexibility enables CAM to interact with several proteins [45]. Conformational changes of CAM have been observed on time scales spanning many orders of magnitude, from rapid wobbling motions (ns timescale) to large transitions from compact to extended forms requiring at least 100 μs [46]. Thus, CAM was selected as a test case for our study because it is considered a challenging system. Indeed, previous long standard MD simulations and other NM-based methods (Clust-ENM) failed to describe all the variability observed in experimental structures [15, 47]

The same protocol adopted for lysozyme was followed: independent NM calculations using distinct models were initially carried out from the equilibrated CAM/Ca^{2+} extended structure in explicit solvent. The subspaces spanned by the five low-frequency NMs are nearly identical to all-atom NM data, as given by the markedly high RMSIPs (0.95 for AAENM and 0.96 for HCA) (Supplementary Figure 3). Hence, we can assume that exploring these subspaces will result in similar conformational populations. Accordingly, we decided to perform a single set of MDeNM simulations for CAM/Ca^{2+}, adopting the simplest HCA model to provide the excitation directions.

The similarity between NMs and PCs calculated from 532 experimental structures was evaluated. Compared to lysozyme results (Figure 3), higher overlaps were obtained, revealing that NMs capture the overall variations observed in the experimental data set. The RMSIP between the subspaces defined by the two sets of vectors was 0.76 (Figure 5A). Figure 5B shows the distributions of the radius of gyration obtained for distinct simulation data sets. While a standard 1 μs MD resulted in a unimodal distribution centered on the starting extended structure, excited simulations using either PCs or NMs resulted in larger distributions. The PCA-based exploration resulted in a smaller sampling than simulations using NMs as excitation directions. All methods failed to reproduce the entire conformational sampling

retrieved from experimental data, mainly composed of compact structures. Although NM-based exploration favored elongated structures, it also covered more compact states (Figure 5B). Therefore, the approach captured this CAM feature better than others tested in this study.

Next, the conformational variability observed from simulations and experimental data was further detailed. To this end, two structural descriptors previously adopted in other studies were chosen [47]. They account for torsions between the two lobes and a bending angle to describe the approximation between them (Figure 2B). The distribution of experimental conformers was split to highlight the impact of CAM binding partners on the protein conformation. The bound structures were clustered around low values of the bending and torsional angles, showing a strong preference for more compact states. On the other hand, the free states were scattered over a large plot area, thus confirming the high structural plasticity of the protein (Figure 5C).

The standard MD simulation covered a limited portion of the conformational space compared to the hybrid approaches. Sampling was uniformly spread around the initial extended structure (yellow square on the plots). The PCA-based exploration resulted in a more extensive sampling range along the bending angle but not along the torsional angle. Interestingly, the NM-based sampling resulted in more extensive sampling regarding both descriptors, covering a significant part of the subspace spanned by the free structures. However, neither method could generate conformers close to the cluster of bound structures (Figure 5C).

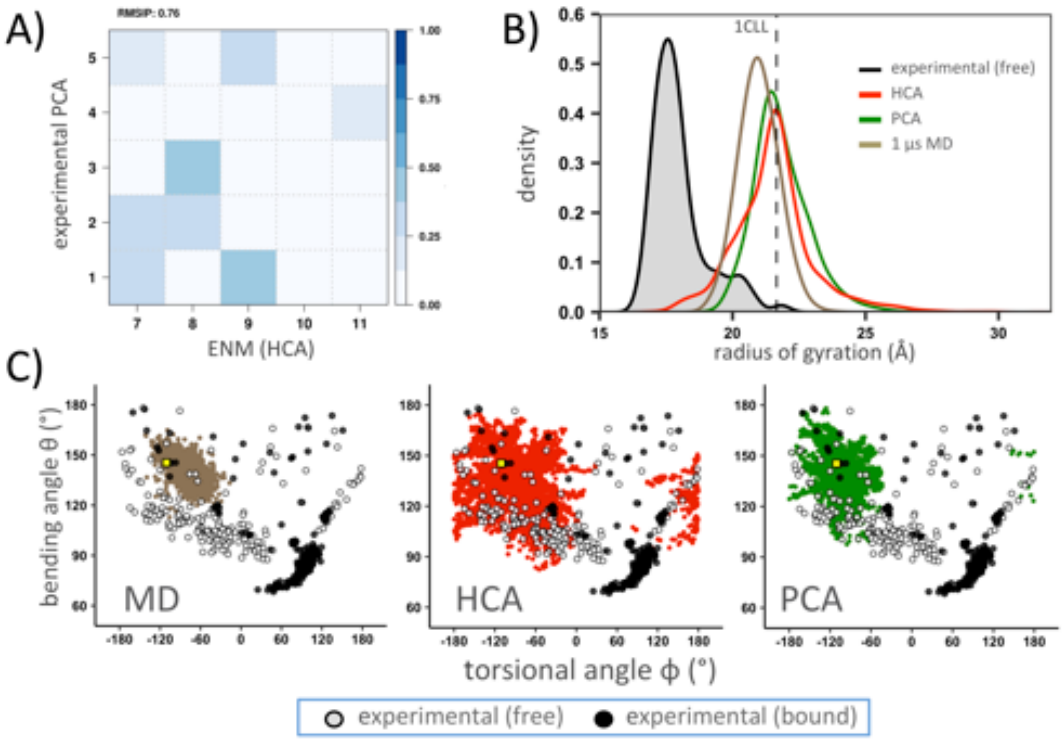

Figure 5: Exploring calmodulin conformational space using MDexciteR. (A) Overlap between low-frequency NMs and PCs describing structural variations in the experimental ensemble. (B) Distributions of the radius of gyration (Rg) calculated from the C_α atoms retrieved from different data sets, as indicated in the legend. The dashed line corresponds to the Rg of the starting structure (PDB ID: 1CLL) (C) Bidimensional plots showing the sampling of calmodulin conformational space using the bending and torsional angles as descriptors. Each simulation set is indicated on the bottom left of the plots. The projection of the starting structure onto this subspace is shown as a yellow square. The projections of experimental structures are also shown as indicated in the legend.

2.3 Test case 3: Beta-2 adrenergic receptor

The β2-adrenergic receptor (β2AR) belongs to the rhodopsin family/class A of G protein-coupled receptors (GPCRs). β2AR was the second GPCR to have its atomic structure unraveled [48, 49]. The data accumulated over the last decade enabled a deep understanding of β2AR activation mechanisms and interactions with ligands [50]. This protein was selected as a test case as it offers an opportunity to evaluate the performance of NM-based exploration and the flexibility of our approach in simulations of a protein-membrane system. Previous studies have provided evidence that low frequency NMs successfully capture diverse aspects of the functional dynamics of GPCRs [51, 52]. A caveat revealed in these studies was the poor description of the flexibility of extra and intracellular segments connecting the transmembrane

helices Indeed, these highly flexible regions move as rigid blocks according to low-frequency NMs [52]. Therefore, the flexibility of connecting segments needs to be better represented in hybrid simulations using low frequency NMs calculated for β2AR (this should also be the case for other GPCRs).

While in the original MDeNM method the excitation was applied to all protein atoms, here we can easily excite a subset of them. Therefore, additional excitation velocities were added only at residues within the transmembrane helices, leaving the connecting segments with more conformational freedom (Supplementary Figure 4). A key point regarding the protocol is the requirement of using low excitation energies to ensure a proper relaxation of these segments during the exploration along each combination of modes.

A significant advantage of MDexciteR is the possibility to work with simplified ENMs and easily explore the same set of motions for the *apo* (receptor only) and *holo* (receptor + agonist) systems (Figure 6A). To this end, NM calculations using the C_α-based HCA model were performed only in the *apo* system and the obtained lowest frequency modes were stored to be used as inputs for the *holo* system. The choice of the *apo* structure for NM calculations was based on the fact that ligand binding generally restricts motions once energetically available in the free state. Based on this strategy, we could directly assess the ligand related effects on the distribution of conformations observed after the simulations.

We inspected the conformational variability obtained after MDeNM simulations using the same protocol for both systems. The starting conformations are different due to the requirement of an extensive equilibration protocol carried out for each system to properly accommodate lipids and water molecules around the protein (see details in the methods section). The distribution obtained for the *apo* form is broader and covers the entire range of helix-3-helix6 distances sampled by the *holo* form. While the narrower unimodal distribution obtained for the *holo* form is centered on the starting structure, the populations obtained in the *apo* form are shifted towards the peak obtained for the *holo* system. This finding suggests that even in the absence of ligands, the intermediate state obtained experimentally in the presence of the agonist BI167107 would be energetically favorable. However, as the *apo* system considered in this study was obtained artificially by removing the ligand from the experimental structure, we cannot exclude the impact of this step on the resulting populations (Figure 6B).

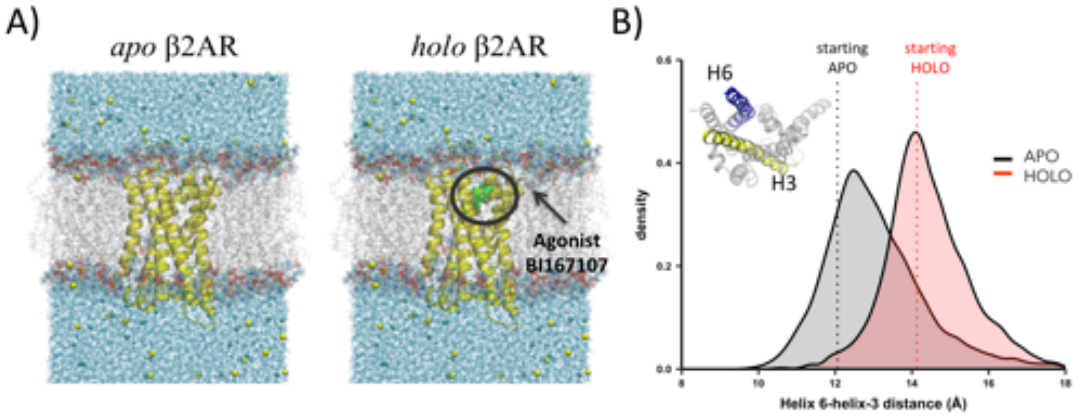

Figure 6: β2AR MDeNM simulations**.** (A) 3D view of the simulated systems considering a realistic environment. The location of the agonist BI167107 is highlighted. (B) Conformational populations obtained after excited simulations. The distributions were calculated from conformers collected after each cycle of excitation/relaxation. The initial helix-3-helix-6 distances are indicated by vertical dotted lines. The inset displays a 3D view of the considered helices.

Finally, we attempted to analyze whether MDeNM simulations would reveal putative β2AR activation motions without ligands. GPCRs are known to exhibit constitutive basal activity even in the absence of agonists [53] . Indeed, the structural rationalization of such mechanism is a topic of great interest. To this end, we compared the displacements obtained from our simulations to the structural variation observed between the initial structure and the active G-protein bound state (PDB ID: 3SN6) [54]. The replicate leading to the largest variations on the helix-3-helix-6 distances was selected for visual comparison here. Our results show that picosecond timescale MDeNM simulations yield intermediate states that comply with the β2AR activation process (Figure 7). These structures may be further relaxed using standard MD or be taken as starting points for simulations using other methods. The generation of such states along the transition towards the active state would require standard MD simulations at the microsecond timescale onwards.

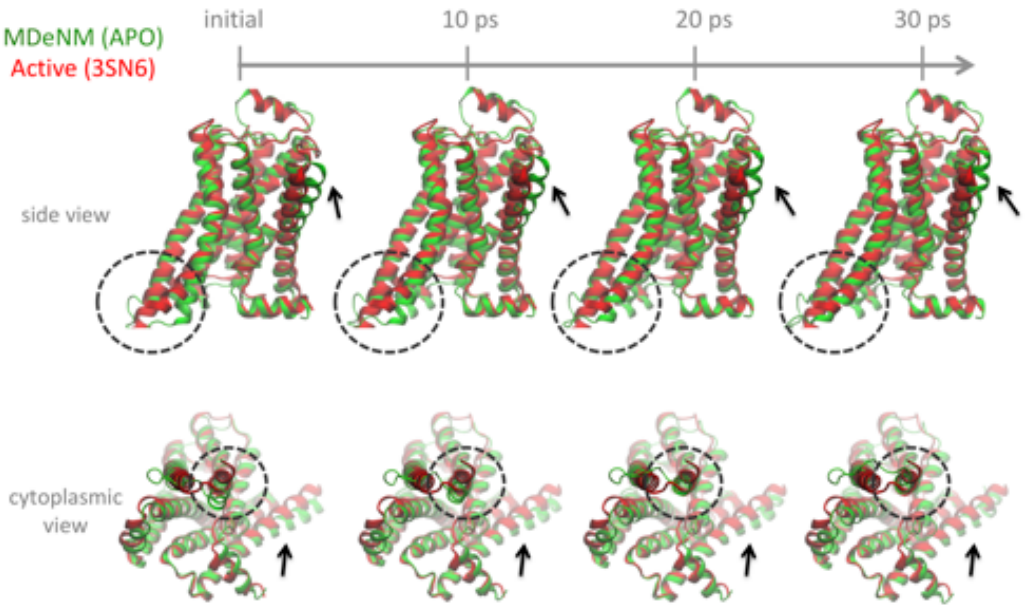

Figure 7: Visual inspection of a β2AR MDeNM trajectory describing receptor activation motions even in the absence of agonists. Snapshots were collected at 10 ps intervals. The superposition with the active conformer bound to Gs protein (PDB ID: 3SN6) [54] is given for comparison purposes. Dotted circles highlight the cytoplasmic outward of helix-6, which is related to β2AR activation. Arrows show the approximation of helix-1to the active conformation obtained experimentally.

2.4 Free energy landscape analysis

Figure 8 shows the free energy landscapes (FEL) obtained from the concatenated relaxed states. Supplementary Figure 5 depicts a schematic view of the protocol (details in the Methods section). The FELs obtained for lysozyme were calculated considering the projections of the trajectories onto normal modes describing the dominant hinge bending and twisting motions (modes 7 and 8, respectively). Clearly, the FELs obtained from MDexciteR (NM and PCA) resemble those generated from a 1 μ s standard MD. In all systems, a dominant population was found in a large basin comprising the entire pool of experimental structures. However, the distribution obtained from MDexciteR (NM) was the largest. It reveals a minor subpopulation corresponding to a more open conformation than those observed in the experimental ensemble (negative values along mode 7).

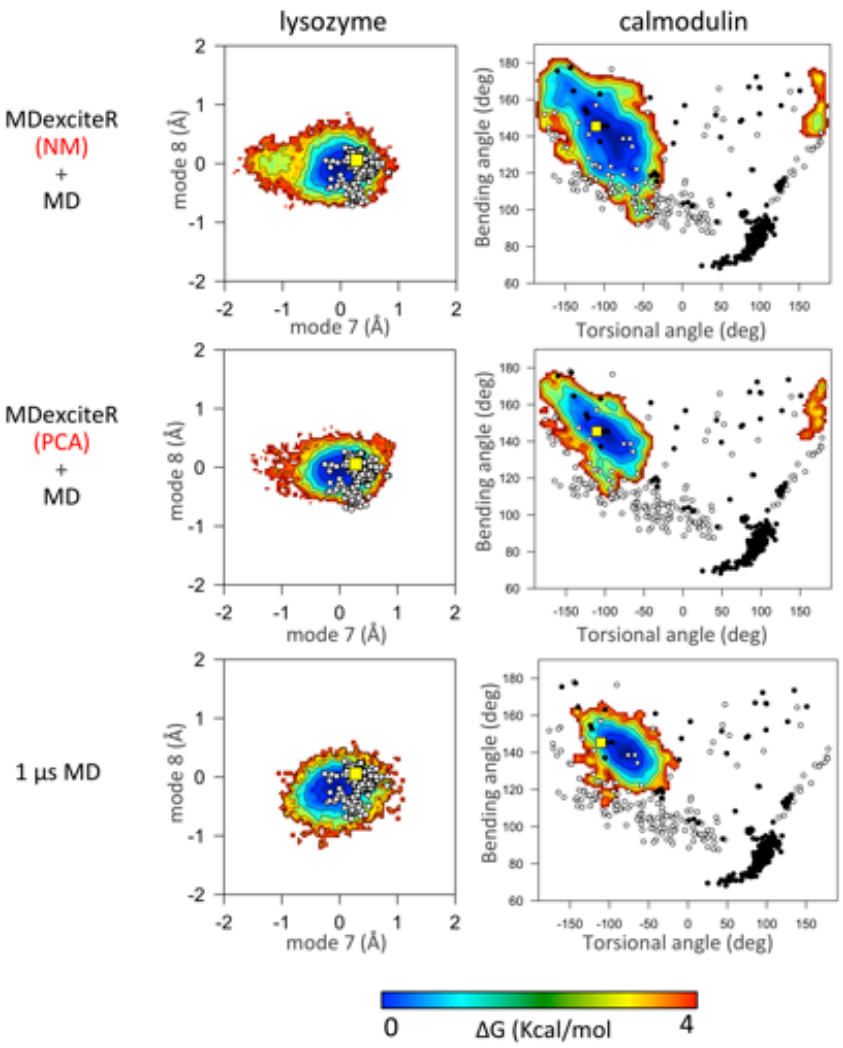

Figure 8: Free energy landscape (FEL) analysis. Lysozyme and calmodulin results are presented in the left- and right-hand side of the picture, respectively. The projections of lysozyme experimental structures are shown as white circles. Experimental structures of CAM in complex with other proteins or ligands are represented as black circles, while free states are colored white. For both test cases, the projection of the starting structure onto each subspace is shown as a yellow square.

Regarding CAM, the sampling obtained from the protocol using MDexciteR was more extensive than those derived from a 1 μs MD. The dominant low energy populations for both NM and PCA-based protocols revealed from extended (bending angle > 130°) to moderately collapsed states (values between 100 and 120°). However, sampling of torsions between CAM lobes was not so efficient. Comparing each data source, we notice a higher sampling efficiency of the protocol combining MDexciteR (NM) even after relaxation. On the other hand, the 1 μs standard MD

simulation generated conformers close to the initial state, with a much narrower distribution compared to the protocols including excited MD (Figure 8).

3. Discussion

In this study we evaluated different aspects of the enhanced sampling using coarse-grained normal modes or principal components. In this section we will discuss each key point and provide guidelines for future studies.

3.1 Choosing relevant NM or PC directions

The choice of specific PCs or NMs for defining the essential subspace is a crucial step. Instead of an arbitrary definition, we decided to use the fractional contribution of the PCs for the overall variance (or total fluctuations in the case of NMs) as the criterion for defining the relevant subspace (Supplementary Figure 2). Once experimental structural data is available, one can adopt other criteria. It may include a subset of motions known to explain a given phenomenon (*i.e.,* active-site opening, interdomain-motions, channel-gating). However, the more directions explored, the higher the complexity of the conformational space to be sampled, therefore requiring a large number of replicas.

3.2 Coarse-grained vs. all atom normal modes as excitation directions

It is well known that NMs calculated with simplified models reproduce well the motions calculated with all-atom models. This feature is evident when evaluating the low-frequency end of the vibrational spectrum, associated with modes describing collective motions [3, 55]. This observation motivated us to test simplified NMs as excitation directions in our MDeNM approach (Supplementary Figures 1 and 3). Interestingly, the conformational populations obtained for lysozyme using NMs calculated with different levels of coarseness are nearly indistinguishable (Figure 4). According to our data, the loss of precision on NMs with simplified models is less relevant under the MDeNM approach. This feature is likely due to the coupling of the excited large amplitude motions with the fast degrees of freedom simulated under explicit solvent effects. Rather than evaluating the best energy function for NM calculations [56], MDeNM focuses on generating energetic relevant conformational

states under realistic conditions, thus yielding highly valuable outputs for various applications.

The adoption of coarse-grained NMs in the MDeNM procedure has significant advantages. Calculations with the ENM representations are faster and less memory-intensive, which is interesting for simulating large systems, such as viral capsids. Furthermore, the energy minimization step in vacuum required for NM calculations using all-atom energy functions is no longer necessary. These features allowed us to introduce NM calculations directly into the MDexciteR implementation developed for this study. In this R script, all input/output processing, NM calculations, and submission of excited MD simulations (using an external program with GPU support, i.e. NAMD and CHARMM) are accomplished from a single user-prepared input (Supplementary text). We believe that such implementation will help the MDeNM approach to reach a wider community. Future studies will include using coarse-grained NMs in excited simulations in conjunction with more sophisticated energy functions, such as polarizable force fields.

3.3 Advantages and limitations of guiding the conformational exploration along experimental PCs

Both NM and PCA-based exploration outperformed standard MD simulations concerning the sampling efficiency. Although the cumulative simulation time for each enhanced sampling approach was 12 ns (100 replicates of 120 ps each), the conformational space sampled was more extensive than standard simulations one hundred times longer. Both strategies performed almost equally well for lysozyme. Regarding this test case, it is crucial to keep in mind that the selection of excitation directions will impact the quality of the results. For example, the widely studied hinge bending motion described by the lowest frequency NM (regardless of the model considered) was the 5th most relevant variation according to PCA analysis (Figure 3). Should our PCA-based exploration be conducted considering the top three vectors that most contribute to the overall variance (thus excluding the 5th vector), the resulting conformational populations would not reflect the dominant contribution of the hinge bending motion.

The NM-based sampling of CAM/Ca^{2+} resulted in a substantially larger conformational space explored. We wondered about the reasons behind this observation. From the inspection of NM and PC vectors, one would expect a similar

sampling performance from the excited simulations. The differences observed might be due to the nature of motions described by low-frequency NMs. Indeed, they are also known as "soft modes", describing intrinsic physical motions with minimum energetic costs [55]. Therefore, following such directions in the MDeNM procedure using low excitation energies would rarely result in the obtaining of non-physical states.

On the other hand, PCs represent a linear decomposition of the structural variations observed in the experimental dataset. Consequently, this procedure could result in non-physical motions associated with high energetic costs [18]. To overcome this drawback, we suggest an initial search for NMs close to the motions of interest encoded in experimental ensembles and subsequently using these modes as excitation directions, instead of directly using the principal components vectors. According to this strategy, the protein dynamics would be driven along intrinsic relaxed motions.

Another issue associated with PCA calculations is the requirement of a prior structural superposition of the conformers included in the ensemble. Indeed, each structure is fitted to the average coordinates. While this rigid-body alignment is very efficient for globular proteins, it may overestimate the displacements of the smaller domains for proteins with multiple domains connected by linkers. This limitation will impact the quality of the PCs, thus increasing the probability of obtaining non-physical structural variations. The dihedral PCA technique can alleviate this issue, but still there is no guarantee that the resulting variations will correspond to actual intrinsic motions.

The simulations using calmodulin as a test case also revealed some limitations of both approaches. Although the excitation of NMs alleviates the impact of the initial conformation on sampling, we still observed a predominance of multiple extended and not fully compact conformations. This result is partially consistent with NMR and FRET data showing that CAM/Ca^{2+} can adopt multiple stable conformers in solution [46, 57]. Future MDeNM simulations of the apo-CAM probably will likely generate an ensemble of more compact states, as revealed by experiments [58]. Considering that local unfolding of CAM central helix is observed in the jumps between extended and compact forms, a promising strategy to improve sampling would be to include higher frequency modes among the excitation directions. According to previous data, these modes are related to protein stability rather than describing slow collective dynamics [59]. Therefore, future studies should test MDeNM protocols including a wider range of

excited NM frequencies. Finally, we believe that full exploration of the calmodulin conformational space would only be achieved by performing simulations under distinct conditions (apo, bound with partners, etc).

3.4 New perspectives opened by the MDexciteR implementation

Besides the advantages of including CG-NMA excitation directions discussed above, the generalization of the MDeNM method proposed here brings other perspectives. First, the simulations conducted for β2AR on this study reveal another key advantage: excitation effects can be applied to a selection of residues, leaving regions whose dynamics is poorly described by low frequency NMs with full conformational freedom. We believe that leaving these regions free is a simple solution perfectly aligned with one of the main motivations behind MDeNM: the idea of enhancing the conformational exploration using minimal biases (Supplementary Figure 4).

Another key advantage is the possibility of using the same set of motions to compare the impact of ligand binding on the steady-state populations. Here, MDeNM simulations on apo β2AR revealed sampling of putative activation motions, leading to a small population that resembles the G-protein bound state (Figures 6 and 7). Rather than a single pathway, several types of motions could contribute to reaching the active structure. However, it is important to notice that helix-1 did not fully transit to the active state. This observation could be associated to the complexity of the β2AR activation mechanism that seems not to be fully described by a single linear combination of NMs. As previously discussed for calmodulin, the re-calculation of modes in intermediate states combined with the selection of new combinations may be required to obtain a complete description of the activation process. The great advantage of our procedure is that no endpoints were required to produce these results. Given that many different pathways can be accessed to reach the G-protein bound state, MDeNM constitutes a powerful tool to rapidly generate a set of putative activation motions that can be further deeply characterized using other methods, such as umbrella sampling or metadynamics.

Regarding the efficiency of the method, we believe that MDeNM based sampling needs further improvement. Two different strategies could be adopted for this purpose. The first would be an initial generation of a pool of structures, followed

by clusterization and selection of a few relevant states. Then, a new round of MDeNM simulations starting from each of these structures would generate conformers far from them. This procedure would ensure that the excitation directions take into account the rearrangements observed in intermediate structures.

Another strategy that follows the same principle consists of recalculating NMs during the time-course of the simulations. This procedure has proven effective for the Cryo-EM fitting procedure based on MDeNM simulations [29]. Adopting these new strategies will likely improve the characterization of complex conformational transitions not well described by a single linear combination of vectors.

3.5 Free energy landscapes characterization

Previous studies demonstrated the efficiency and accuracy of our protocol for calculating FELs from MDeNM simulations, in agreement with experimental data, metadynamics, and umbrella sampling calculations[14, 17]. Its advantage is the simplicity, still leading to accurate predictions for several applications. Both excited MD strategies lead to a more extensive range of sampling than orders of magnitude longer unbiased MD (Figure 8). This observation is more evident for a highly flexible system such as calmodulin. However, after relaxation, the sampling differences between each exciting MD strategy are less noticeable than during the excitation phase (Figure 5C). This expected effect comes from a quick relaxation of non-favorable excited states. Still, NM-based sampling remains the most efficient approach.

On the other hand, the distributions obtained for lysozyme were very similar, regardless of the adopted enhanced sampling strategy. This protein is considerably more rigid than calmodulin, as given by the high structural similarity observed in experimental structures. The slight opening of the system only noticed using the NM-based sampling is likely relevant for ligand entrance / product release, as previously revealed [60]. It is evident that our protocol (consisting of less than one hundred nanosecond-long MD simulations) was enough to cover all variability observed experimentally.

4. Conclusion

NM-based methods have been demonstrating high efficiency for characterizing protein conformational variability without compromising their accuracy [33, 55]. MDeNM stands out from other methodologies because the enhanced sampling is achieved with minimal interference, and the simulations are performed with the same level of detail as standard MD [29]. In addition, velocities injected along the intrinsic motions described by NMs constitute an almost negligible increase in the overall kinetic energy of the system. Moreover, the additional energy is quickly dissipated, enabling the combination of MDeNM and standard simulations to determine free energy landscapes [29].

Previous MDeNM results reported by other studies were obtained using all-atom NMs calculated using standard energy functions [14, 28, 30, 32, 33, 61]. The success of methods based on simplified NMs motivated us to develop a new MDeNM implementation allowing the selection of such modes as excitation directions. This effort enabled us to assess whether the method's accuracy would be preserved using NMs calculated from simplified forcefields, or PCs extracted from experimental ensembles.

This study provides strong evidence that NMs calculated from elastic network representations can be used in MDeNM simulations without sacrificing the method's accuracy. On the other hand, some pitfalls were revealed regarding the PCA-based exploration of conformational spaces. First, the sampling achieved using the first PC vectors (ranked by their contributions to the overall variance) is substantially less efficient than NM-based exploration. Second, the results obtained for calmodulin reveal that even though the excited subspaces were similar, NM-based exploration better agrees with the experimental data. Further, PCs are not guaranteed to describe actual motions from a physical point of view, so we discussed an alternative approach using NMs that are close to the experimental variations.

Finally, the new implementation allows the integration with different simulation programs. We expect our method to reach a larger community based on this feature. In addition, the simplicity of working with ENMs and the minimal requirements for their computation will also contribute to MDeNM popularity. This study also paves the way for developing new protocols based on the re-calculation of NMs on the time course of the simulations. This approach will now be possible thanks

to the lack of necessary energy minimizations required prior to NM calculations using ENMs. Further studies will be conducted to develop and validate new sampling strategies based on MDeNM.

5. Methods

Atomic coordinates and experimental data sets

The following atomic coordinates were selected as starting structures for the calculations: lysozyme (PDB ID: 2LZT) [62] , CAM/Ca^{2+} bound (PDB ID: 1CLL) [63], and β2AR (PDB ID: 3P0G) [64]. Alternative experimental conformations of lysozyme and calmodulin were obtained using ProDy v.1.9.3 [65]. To this end, we carried out a blast search against the Protein Data Bank to retrieve structures sharing at least 90% sequence identity with the reference structures. Subsequently, the coordinates of C_α atoms of these structures were superimposed using the Kabsch's algorithm. A list of the PDB IDs included in each experimental data set is provided in the Supplementary Material. Only the coordinates of C_α atoms were taken into account for subsequent PCA calculations and comparisons with NM vectors.

The following definitions were adopted for the structural descriptors considered in this article. The lysozyme breathing angle was defined from the computation of the centers of mass (COM) of C_α atoms from three regions: (1) residues 28-31 and 111-114; (2) 90-93 and (3) 44-45 and 51-52 [66]. Calmodulin bending and torsional angles were defined from selected points as previously defined [47]. The (COMs) of the N-lobe, C-lobe and the linker defined the bending angle θ. Residues 69 and 91 were selected to represent the linker's initial and final positions. Four points in the space define the virtual dihedral angle φ: the COMs of each lobe and the initial and final positions of the linker (represented by residues 69 and 91, respectively). The helix-3-helix-6 distance adopted to evaluate β2AR was defined by the positions of C_α atoms of residues 131 and 272. All molecular graphics were produced with VMD v. 1.9.3 [67].

Preparation of molecular systems for subsequent simulations

Preparation steps were conducted with the CHARMM-GUI webserver [68]. A similar protocol was adopted for lysozyme and calmodulin, while specific steps were required for β2AR. Unless otherwise specified, the following procedures were

common to all simulated systems. Hydrogen atoms were added to the crystal structures using the HBUILD routine. The simulations were carried out in explicit solvent using periodic boundary conditions. Van der Waals interactions were calculated up to 10 Å, being approximated until 12 Å by using a switching function. Electrostatic interactions were treated with the PME algorithm using a 10-Å cut-off [69]. The SETTLE [70] and SHAKE [71] algorithms were used during MD simulations to fix bonds involving hydrogen atoms in water molecules and protein. Pressure was kept constant at 1 atm during equilibration and production using the Langevin piston method [72]. In these steps, temperature was also kept constant at 300 K using the Langevin thermostat with a damping coefficient of 1 ps^{-1}.

Preparation protocol for lysozyme and calmodulin: each system was placed in a cubic box with a 14 Å layer of TIP3P water molecules [73] (approximately 25.000, depending on the system) added around the solute molecules. Counter-ions were inserted to neutralize the system reaching a 0.15 M NaCl concentration. Then, an energy-minimization protocol was performed, starting by the conjugate-gradient algorithm, keeping protein heavy atoms harmonically restrained with a force constant of 50 kcal mol^{-1} $Å^{-2}$ to avoid structural distortions. The following steps using the same algorithm were carried out using decreasing force constants (up to 2.5 kcal mol^{-1} $Å^{-2}$). Then the atomic velocities were assigned accordingly to a Maxwell−Boltzmann distribution corresponding to 50 K and then slowly increased to 300 K during a 1 ns heating MD using a 1fs integration time. In the equilibration step, the positional restraints were gradually decreased to zero during the first half of a 3 ns constant temperature MD, while in the remaining part, all restraints were removed. The MD integration time adopted for equilibration and production simulations was 2 fs. Production simulations were performed in the NPT ensemble using the same parameters for temperature and pressure control.

Preparation protocol for β2AR: the CHARMM-GUI membrane builder module was used to assemble the protein/membrane/water/ions system [74]. The protein was inserted into a homogenous POPC bilayer. An excess of ~15.000 TIP3 water molecules were added to solvate the protein/membrane system. Counter ions were added to neutralize the system, reaching a 0.15 M NaCl concentration. The energy minimization protocol was performed following the same steps used for the other proteins considered in this study. A four-step equilibration was required here due to the increased complexity of the system. Initially, a short 100 ps equilibration MD in

the NVT ensemble was carried out keeping the protein fixed. Then a similar simulation in the NPT ensemble was performed. In these steps a 1 fs integration time was adopted. In the third step, the protein backbone was restrained with a force constant of 100 kcal mol^{-1} $Å^{-2}$ during a 50 ps MD. The restraints were then gradually decreased during subsequent 50 ps MD. The final equilibration step consisted of a 100 ns MD without any restraints to accommodate lipid and water molecules properly. In the latter steps, a 2 fs integration step was adopted.

All-atom Normal mode calculations using a physical-based forcefield

The 200 lowest-frequency normal modes of lysozyme and calmodulin were calculated with the VIBRAN module of the available academic version of CHARMM (v.42b2) [38]. Calculations were performed on the structures obtained after the equilibration step. Solvent and counter-ions were removed, leaving only the solute. The four Ca^{2+} ions bound to calmodulin were considered in NM calculations. Electrostatic interactions were treated using a distance-dependent dielectric constant ($\varepsilon = 2r_{i,j}$). Prior to NM calculations, an energy minimization protocol was applied, using different algorithms: steepest descent (SD), conjugate-gradient (CG) and the Adopted Basis Newton-Raphson (ABNR). First, harmonic restraints were applied to hold backbone atoms close to their initial positions during the SD steps. The restraints were progressively decreased from 250 kcal mol^{-1} $Å^{-2}$ to zero. Subsequently, additional steps of energy minimization were performed with the CG and the ABNR algorithms without positional restraints, until they reached a convergence criterion of 10^{-5} kcal mol-1 $Å^{-1}$ energy gradient.

Normal mode calculations using simplified models

Normal modes calculations were performed using the R software in conjunction with the *nma.pdb* function included in the Bio3D package [35]. Calculations were performed directly on the equilibrated structures, without the requirement of energy minimization steps. Two different models were considered independently. Both the Harmonic C_α potential (HCA) and the AAENM models are based on elastic networks with distance-dependent force constants (as previously discussed in reference [56]). While the HCA model considers only C_α atoms for the construction of the elastic network, the AAENM model considers all protein heavy atoms. The

rotation-translation blocks (RTB) approximation [75] (one residue per rigid block) was adopted for calculations with the AAENM model.

Principal component analysis of experimental structures

The most relevant structural variations found in each dataset of experimental structures were characterized with principal component analysis (PCA). Briefly, PCA is based on the diagonalization of the covariance matrix, $C_{(i,j)}$, of atomic coordinates displacements whose elements are represented by

$$C_{(i,j)} = \langle \Delta r_i \cdot \Delta r_j \rangle$$

where Δr_i and Δr_j are the displacement vectors of atoms i and j, respectively, from their average positions. Brackets indicate ensemble averages. Then, an eigenvalue problem is solved, resulting in 3N PCs that can be sorted according to their fractional contributions to the overall variance. These calculations were performed using the R software in conjunction with the *pca.xyz* function of the Bio3D library [35].

Molecular dynamics with excited NMs or PCs

All simulations followed the flowchart presented in Figure 1. Apart from the MD simulations using excitation velocities, all other steps were conducted using R v.3.5.1 [36] and the Bio3D package v. 2.3.3 [34, 35]. MD simulations were carried out using NAMD 2.13 [37]. We performed a set of independent simulations following the same protocol for each set of vectors considered. Therefore, direct comparisons between the resulting conformers are allowed.

MDeNM protocol for lysozyme and calmodulin: we performed one hundred independent simulations driving the protein along a distinct linear combination of NM or PC vectors. The top five vectors contributing the most overall fluctuations (in the case of NMs) or the variance (in the case of PCs) define the essential subspace. According to Supplementary Figure 2, the cumulative contribution provided by the first five PCs accounted for more than 75 % of overall variance, while the respective normal modes contributed to more than 75 % of total fluctuations.

Each replica consisted of 30 short excitation/relaxation cycles. The simulation length of each cycle was 5 ps. The maximum excitation temperature adopted was 4 K. Therefore, each independent replicate accounted for 120 ps simulation length. The conformers generated at the end of each cycle were retrieved for further analysis.

MDeNM protocol for β2AR: Five normal mode vectors obtained from NM calculations on the *apo β2AR* using the HCA model defined the essential subspace. These vectors were linearly combined (as described in ref [14] and in the Supplementary Text) producing one hundred combinations of vectors that were stored (*.crd* files). These combinations were used as inputs for the MDeNM simulations of the *holo β2AR* system. Excitation energies were only applied to residues located at transmembrane helices (residues: 29 to 61; 66 to 97; 102 to 137; 146 to 172; 197 to 228; 266 to 299 and 304 to 329). Low excitation temperatures were adopted (maximum T_{nm} = 3 K), therefore ensuring a slower exploration of collective motions and enabling a better conformational adjustment of loop segments. The number of replicas, cycles and the simulation lengths used here were the same as those adopted for the other test cases.

Free energy landscape analysis

Overall, the protocol for the free energy landscape calculation consisted of a clustering step from the concatenated structures generated by MDexciteR followed by relaxation MD simulations starting from each centroid. The gromos algorithm was employed using an RMS cutoff of 1 Å. Supplementary Table 1 summarizes the number of centroids selected for each state. Briefly, from each centroid, velocities were re-generated at 300 K and the backbone and side chain atoms were restrained with a force constant of 5 and 2.5 $kcal\ mol^{-1}\ Å^{-2}$, respectively. These restraints were progressively decreased to zero over the course of a 500 *ps* MD in the NPT ensemble using the same parameters adopted in previous steps. From this point, an unrestrained 1 ns production MD simulation was conducted. For each considered data set, the generated trajectories (recorded at 2 *ps* intervals) were concatenated into a single file considered for further analysis.

We used an in-house R script to calculate the free energy landscapes. 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:

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

where k_B is the Boltzmann constant, T is the temperature of the simulations (300 K), and $P(q_\alpha)$ is an estimate of the probability density function obtained from bi-

dimensional kernel density estimates of distinct properties of interest *q (i.e.,* torsion and bending angles, projections onto NM vectors) calculated over the concatenated standard MD trajectories starting from centroids. $P_{max}(q)$ is the probability of the most visited state.

Overlap between NM vectors and RMISP of essential subspaces

The overlap between a given mode vector, **M**, and another mode vector **X**, is evaluated by their normalized projection:

$$O_i(X) = \frac{\boldsymbol{M_i X}}{\|\boldsymbol{M}_i\|\|\boldsymbol{X}\|}$$

In addition, we measured the similarity between the subspaces defined by the five lowest frequency NMs or the top five PCs contributing to overall variance. To this end, we calculated the RMS inner product (RMSIP). Only displacements of C_α atoms were considered in our comparisons. All overlap and RMSIP calculations were carried out with R/bio3D using the *rmsip* function [35].

Associated Content

Figure S1: Overlap between lysozyme low-frequency NMs. Figure S2: Cumulative contribution of NMs and PCs for overall variance. Figure S3: Overlap between calmodulin low-frequency NMs. Figure S4: Testing the implementation on a G-protein coupled receptor. Figure S5: Schematic representation of the protocol utilized for free energy landscape analysis. Figure S6: Extending the motions of Cα atoms to other selections. Table S1: Details of the simulations for free energy landscape analysis. Additional details of the MDexciteR implementation as well as lists of PDB identifiers of the lysozyme and calmodulin structures considered for comparisons with simulated data. Example files and a tutorial for testing MDexciteR are provided in the following link: https://github.com/mcosta27/MDexciteR

Author Information

Corresponding authors: Mauricio G.S. Costa & David Perahia, e-mail addresses: mauricio.costa@fiocruz.br; david.perahia@ens-paris-saclay.fr.

Author Contributions: MGSC: designed and performed research, analyzed data and wrote the paper. PRB: performed research, analyzed data and reviewed the paper. AG: performed research and reviewed the paper. LSB: performed research. ML: performed research and reviewed the paper. NF: performed research and reviewed the paper. PMB: analyzed data and reviewed the paper. DP: designed research, analyzed data and wrote the paper. All authors have given approval to the final version of the manuscript.

Note: The authors declare no competing interests.

Acknowledgments

All authors would like to thank the Brazilian-French collaborative project CAPES/COFECUB (project 873/2017). This work was performed using HPC resources from the Mésocentre computing center of CentraleSupélec and ENS Paris-Saclay supported by CNRS and the Région Île-de-France (http://mesocentre.centralesupelec.fr/).

References

1. Hite, R. K.; MacKinnon, R., Structural Titration of Slo2.2, a Na(+)-Dependent K(+) Channel. *Cell* **2017**, *168*, 390-399 e311.
2. Fenwick, R. B.; van den Bedem, H.; Fraser, J. S.; Wright, P. E., Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc Natl Acad Sci U S A* **2014**, *111*, E445-454.
3. Bahar, I.; Cheng, M. H.; Lee, J. Y.; Kaya, C.; Zhang, S., Structure-Encoded Global Motions and Their Role in Mediating Protein-Substrate Interactions. *Biophys J* **2015**, *109*, 1101-1109.
4. Kokkinidis, M.; Glykos, N. M.; Fadouloglou, V. E., Protein flexibility and enzymatic catalysis. *Adv Protein Chem Struct Biol* **2012**, *87*, 181-218.
5. Stoisser, T.; Brunsteiner, M.; Wilson, D. K.; Nidetzky, B., Conformational flexibility related to enzyme activity: evidence for a dynamic active-site gatekeeper function of Tyr(215) in Aerococcus viridans lactate oxidase. *Sci Rep* **2016**, *6*, 27892.
6. Luk, L. Y.; Loveridge, E. J.; Allemann, R. K., Protein motions and dynamic effects in enzyme catalysis. *Phys Chem Chem Phys* **2015**, *17*, 30817-30827.
7. Mahajan, S.; Sanejouand, Y.-H., On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Archives of Biochemistry and Biophysics* **2015**, *567*, 59-65.
8. Gur, M.; Madura, J. D.; Bahar, I., Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase. *Biophys J* **2013**, *105*, 1643-1652.
9. Shao, Q., Enhanced conformational sampling technique provides an energy landscape view of large-scale protein conformational transitions. *Phys Chem Chem Phys* **2016**, *18*, 29170-29182.
10. Orellana, L.; Yoluk, O.; Carrillo, O.; Orozco, M.; Lindahl, E., Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat Commun* **2016**, *7*, 12575.
11. Lopez-Blanco, J. R.; Aliaga, J. I.; Quintana-Orti, E. S.; Chacon, P., iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res* **2014**, *42*, W271-276.
12. Tekpinar, M.; Zheng, W., Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins* **2010**, *78*, 2469-2481.
13. Franklin, J.; Koehl, P.; Doniach, S.; Delarue, M., MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. *Nucleic Acids Res* **2007**, *35*, W477-482.
14. Costa, M. G.; Batista, P. R.; Bisch, P. M.; Perahia, D., Exploring free energy landscapes of large conformational changes: molecular dynamics with excited normal modes. *J Chem Theory Comput* **2015**, *11*, 2755-2767.
15. Kurkcuoglu, Z.; Bahar, I.; Doruker, P., ClustENM: ENM-Based Sampling of Essential Conformational Space at Full Atomic Resolution. *J Chem Theory Comput* **2016**, *12*, 4549-4562.
16. Zhang, Z.; Shi, Y.; Liu, H., Molecular dynamics simulations of peptides and proteins with amplified collective motions. *Biophys J* **2003**, *84*, 3583-3593.

17. 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.
18. Amadei, A.; Linssen, A. B.; Berendsen, H. J., Essential dynamics of proteins. *Proteins* **1993**, *17*, 412-425.
19. Amadei, A.; Ceruso, M. A.; Di Nola, A., On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins* **1999**, *36*, 419-424.
20. Michielssens, S.; van Erp, T. S.; Kutzner, C.; Ceulemans, A.; de Groot, B. L., Molecular dynamics in principal component space. *J Phys Chem B* **2012**, *116*, 8350-8354.
21. Peng, J.; Zhang, Z., Simulating Large-Scale Conformational Changes of Proteins by Accelerating Collective Motions Obtained from Principal Component Analysis. *J Chem Theory Comput* **2014**, *10*, 3449-3458.
22. Shkurti, A.; Styliari, I. D.; Balasubramanian, V.; Bethune, I.; Pedebos, C.; Jha, S.; Laughton, C. A., CoCo-MD: A Simple and Effective Method for the Enhanced Sampling of Conformational Space. *Journal of Chemical Theory and Computation* **2019**, *15*, 2587-2596.
23. Abrams, C. F.; Vanden-Eijnden, E., Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc Natl Acad Sci U S A* **2010**, *107*, 4961-4966.
24. Chen, M., Collective variable-based enhanced sampling and machine learning. *The European Physical Journal B* **2021**, *94*, 211.
25. Spiwok, V.; Králová, B.; Tvaroška, I., Continuous metadynamics in essential coordinates as a tool for free energy modelling of conformational changes. *Journal of Molecular Modeling* **2008**, *14*, 995-1002.
26. Spiwok, V.; Lipovová, P.; Králová, B., Metadynamics in Essential Coordinates: Free Energy Simulation of Conformational Changes. *The Journal of Physical Chemistry B* **2007**, *111*, 3073-3076.
27. Batista, P. R.; Robert, C. H.; Marechal, J. D.; Hamida-Rebai, M. B.; Pascutti, P. G.; Bisch, P. M.; Perahia, D., Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis--application to the HIV-1 protease. *Phys Chem Chem Phys* **2010**, *12*, 2850-2859.
28. Floquet, N.; Costa, M. G.; Batista, P. R.; Renault, P.; Bisch, P. M.; Raussin, F.; Martinez, J.; Morris, M. C.; Perahia, D., Conformational Equilibrium of CDK/Cyclin Complexes by Molecular Dynamics with Excited Normal Modes. *Biophys J* **2015**, *109*, 1179-1189.
29. Costa, M. G. S.; Fagnen, C.; Venien-Bryan, C.; Perahia, D., A New Strategy for Atomic Flexible Fitting in Cryo-EM Maps by Molecular Dynamics with Excited Normal Modes (MDeNM-EMfit). *J Chem Inf Model* **2020**, *60*, 2419-2423.
30. Dudas, B.; Toth, D.; Perahia, D.; Nicot, A. B.; Balog, E.; Miteva, M. A., Insights into the substrate binding mechanism of SULT1A1 through molecular dynamics with excited normal modes simulations. *Scientific Reports* **2021**, *11*, 13129.
31. Fagnen, C.; Bannwarth, L.; Zuniga, D.; Oubella, I.; De Zorzi, R.; Forest, E.; Scala, R.; Guilbault, S.; Bendahhou, S.; Perahia, D.; Vénien-Bryan, C., Unexpected Gating Behaviour of an Engineered Potassium Channel Kir. *Front Mol Biosci* **2021**, *8*, 691901.

32. Dudas, B.; Perahia, D.; Balog, E., Revealing the activation mechanism of autoinhibited RalF by integrated simulation and experimental approaches. *Scientific Reports* **2021**, *11*, 10059.
33. Kaynak, B. T.; Krieger, J. M.; Dudas, B.; Dahmani, Z. L.; Costa, M. G. S.; Balog, E.; Scott, A. L.; Doruker, P.; Perahia, D.; Bahar, I., Sampling of Protein Conformational Space Using Hybrid Simulations: A Critical Assessment of Recent Methods. *Frontiers in Molecular Biosciences* **2022**, *9*.
34. Grant, B. J.; Rodrigues, A. P.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S., Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695-2696.
35. Skjaerven, L.; Yao, X. Q.; Scarabelli, G.; Grant, B. J., Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **2014**, *15*, 399.
36. R Core Team *R: A language and environment for statistical computing.*, R Foundation for Statistical Computing, Vienna, Austria 2015.
37. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005**, *26*, 1781-1802.
38. 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*, 1545-1614.
39. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., *AMBER 14*. 2014.
40. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435-447.
41. 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*, 4995-4999.
42. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71-73.
43. Yao, X. Q.; Skjaerven, L.; Grant, B. J., Rapid Characterization of Allosteric Networks with Ensemble Normal Mode Analysis. *J Phys Chem B* **2016**, *120*, 8276-8288.
44. Hinsen, K.; Petrescu, A.-J.; Dellerue, S.; Bellissent-Funel, M.-C.; Kneller, G. R., Harmonicity in slow protein dynamics. *Chemical Physics* **2000**, *261*, 25-37.
45. Villalobo, A.; Ishida, H.; Vogel, H. J.; Berchtold, M. W., Calmodulin as a protein linker and a regulator of adaptor/scaffold proteins. *Biochim Biophys Acta Mol Cell Res* **2018**, *1865*, 507-521.

46. Slaughter, B. D.; Allen, M. W.; Unruh, J. R.; Bieber Urbauer, R. J.; Johnson, C. K., Single-Molecule Resonance Energy Transfer and Fluorescence Correlation Spectroscopy of Calmodulin in Solution. *The Journal of Physical Chemistry B* **2004**, *108*, 10388-10397.
47. Aykut, A. O.; Atilgan, A. R.; Atilgan, C., Designing Molecular Dynamics Simulations to Shift Populations of the Conformational States of Calmodulin. *PLOS Computational Biology* **2013**, *9*, e1003366.
48. Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C., High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258-1265.
49. Rasmussen, S. G.; Choi, H. J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F.; Weis, W. I.; Kobilka, B. K., Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **2007**, *450*, 383-387.
50. Wu, Y.; Zeng, L.; Zhao, S., Ligands of Adrenergic Receptors: A Structural Point of View. *Biomolecules* **2021**, *11*.
51. Damian, M.; Mary, S.; Maingot, M.; M'Kadmi, C.; Gagne, D.; Leyris, J. P.; Denoyelle, S.; Gaibelet, G.; Gavara, L.; Garcia de Souza Costa, M.; Perahia, D.; Trinquet, E.; Mouillac, B.; Galandrin, S.; Gales, C.; Fehrentz, J. A.; Floquet, N.; Martinez, J.; Marie, J.; Baneres, J. L., Ghrelin receptor conformational dynamics regulate the transition from a preassembled to an active receptor:Gq complex. *Proc Natl Acad Sci U S A* **2015**, *112*, 1601-1606.
52. Louet, M.; Karakas, E.; Perret, A.; Perahia, D.; Martinez, J.; Floquet, N., Conformational restriction of G-proteins Coupled Receptors (GPCRs) upon complexation to G-proteins: a putative activation mode of GPCRs? *FEBS Lett* **2013**, *587*, 2656-2661.
53. Chidiac, P.; Hebert, T. E.; Valiquette, M.; Dennis, M.; Bouvier, M., Inverse agonist activity of beta-adrenergic antagonists. *Mol Pharmacol* **1994**, *45*, 490-499.
54. Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K., Crystal structure of the β2 adrenergic receptor–Gs protein complex. *Nature* **2011**, *477*, 549-555.
55. Krieger, J. M.; Doruker, P.; Scott, A. L.; Perahia, D.; Bahar, I., Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods. *Current Opinion in Structural Biology* **2020**, *64*, 34-41.
56. Fuglebakk, E.; Reuter, N.; Hinsen, K., Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions. *Journal of Chemical Theory and Computation* **2013**, *9*, 5618-5628.
57. Gsponer, J.; Christodoulou, J.; Cavalli, A.; Bui, J. M.; Richter, B.; Dobson, C. M.; Vendruscolo, M., A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure* **2008**, *16*, 736-746.
58. Zhang, M.; Tanaka, T.; Ikura, M., Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nature Structural Biology* **1995**, *2*, 758-767.

59. Bahar, I.; Atilgan, A. R.; Demirel, M. C.; Erman, B., Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Physical Review Letters* **1998**, *80*, 2733-2736.
60. De Simone, A.; Aprile, F. A.; Dhulesia, A.; Dobson, C. M.; Vendruscolo, M., Structure of a low-population intermediate state in the release of an enzyme product. *Elife* **2015**, *4*.
61. Fagnen, C.; Bannwarth, L.; Oubella, I.; Forest, E.; De Zorzi, R.; de Araujo, A.; Mhoumadi, Y.; Bendahhou, S.; Perahia, D.; Vénien-Bryan, C., New Structural insights into Kir channel gating from molecular simulations, HDX-MS and functional studies. *Scientific Reports* **2020**, *10*, 8392.
62. Ramanadham, M.; Sieker, L. C.; Jensen, L. H., Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares. *Acta Crystallogr B* **1990**, *46 (Pt 1)*, 63-69.
63. Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A., Calmodulin structure refined at 1.7 Å resolution. *Journal of Molecular Biology* **1992**, *228*, 1177-1192.
64. Rasmussen, S. G. F.; Choi, H.-J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; DeVree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; Steyaert, J.; Weis, W. I.; Kobilka, B. K., Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* **2011**, *469*, 175-180.
65. Bakan, A.; Bahar, I., The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc Natl Acad Sci U S A* **2009**, *106*, 14349-14354.
66. 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*, 6480-6486.
67. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14*, 33-38, 27-38.
68. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **2008**, *29*, 1859-1865.
69. 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.
70. Miyamoto, S.; Kollman, P. A., Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of computational chemistry* **1992**, *13*, 952-962.
71. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327-341.
72. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **1995**, *103*, 4613-4621.
73. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926-935.
74. Jo, S.; Kim, T.; Im, W., Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE* **2007**, *2*, e880.

75. Durand, P.; Trinquier, G.; Sanejouand, Y.-H., A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* **1994**, *34*, 759-771.

For table of Contents Graphic only