# Combining Virtual Reality Visualization with Ensemble Molecular Dynamics to Study Complex Protein Conformational Changes

Jordi Juárez-Jiménez[†θ], Philip Tew[$], Michael O´Connor[º‡•], Salome Llabrés[†], Rebecca Sage[$], David Glowacki[º‡•] and Julien Michel[†*]

[†] EaStCHEM School of Chemistry, University of Edinburgh, David Brewster road, Edinburgh EH9 3FJ, United Kingdom.

[$] Interactive Scientific, Engine Shed, Station Approach, Bristol BS1 6QH, United Kingdom.

[º] Intangible Realities Laboratory, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom.

[‡] Dept. of Computer Science, University of Bristol, Merchant Venture's Building, Bristol BS8 1UB, United Kingdom.

[•] Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom.

[θ] Current address: Facultat de Farmacia i Ciencies de l'Alimentació, Universitat de Barcelona, Av. Joan XXIII 27-31, Barcelona, 08028, Spain.

[*] Corresponding author. E-mail: mail@julienmichel.net

**Abstract**

Molecular dynamics (MD) simulations are increasingly used to elucidate relationships between protein structure, dynamics and their biological function. Currently it is extremely challenging to perform MD simulations of large-scale structural rearrangements in proteins that occur on millisecond timescales or beyond, as this requires very significant computational resources, or the use of cumbersome 'collective variable' enhanced sampling protocols. Here we describe a framework that combines ensemble MD simulations and virtual-reality visualization (eMD-VR) to enable users to interactively generate realistic descriptions of large amplitude, millisecond timescale protein conformational changes in proteins. Detailed tests demonstrate that eMD-VR substantially decreases the computational cost of folding simulations of a WW domain, without the need to define collective variables *a priori*. We further show that eMD-VR generated pathways can be combined with Markov State Models to describe the thermodynamics and kinetics of large-scale loop motions in the enzyme cyclophilin A. Our results suggest eMD-VR is a powerful tool for exploring protein energy landscapes in bioengineering efforts.

**Introduction**

There is much interest in the use of molecular dynamics (MD) simulations to both guide the design and interpretation of experiments and elucidate relationships between the sequence, structure, dynamics and function of biological molecules.[1-13] A current frontier for the field is the realistic simulation of large scale motions that occur on millisecond timescales or beyond in proteins.[14,15] Such processes currently require either an enormous amount of computing power that exceeds the capabilities of most supercomputers,[16,17] or the use of enhanced sampling methodologies that accelerate motions along predefined sets of collective variables (CV). CV-based methodologies are particularly useful when the endpoints of the conformational change have been previously structurally characterized, and a realistic interconversion pathway between the endpoints can be expressed with low-dimensionality descriptors suitable for biasing MD simulations via methodologies such as umbrella sampling[18] or metadynamics[19] for instance. These requirements limit applicability to well-chosen problems.[20-37]

Interactive MD simulations (iMD) tackle the problem differently by allowing users to manipulate in real time an MD simulation to achieve the desired conformational change.[38-41] A growing body of literature suggests that iMD simulations enable humans to generate complex conformational changes that may overcome the limitations of CV-based methods. Alongside these developments, the availability of commodity virtual reality (VR) equipment has also led to a surge in the development of VR-based molecular visualization approaches that facilitate an intuitive understanding of the nature of molecular interactions for chemical education and for research purposes.[42-58] Recently, these technologies were combined to enable interactive MD simulations in virtual reality (iMD-VR) for a range of chemical systems.[38,59] Notably, iMD-VR has been shown to be useful in generating plausible binding modes of protein-ligand

complexes, along with the corresponding conformational changes (e.g., loop motions) required to undertake the binding and unbinding processes.[60]

To maintain an effective user experience, current iMD-VR approaches tend to add biasing forces to the MD simulations which cause the system to respond in (near) real time. Current highly optimized MD engines such as OpenMM[61] or AMBER[62] achieve performances of ca. 200-500 ps of MD sampling per minute of real time for a modestly sized protein solvated in a box (ca. 25k atoms). Such sampling time is several orders of magnitude below the timescale of protein folding or large-scale conformational changes, which typically occur on microseconds or slower. Thus it may be necessary to add strong biasing forces in an iMD-VR simulation to force a conformational change to happen over a timescale of a few seconds to minutes of real time. An open question is whether the work done on the system by a strong bias results in the generation of pathways that deviate significantly from the minimum free energy path.[63] The issue is likely to be more problematic for investigations of large scale biomolecular assemblies that may include millions of particles.[64-69] Additionally, the interactive exploration of conformational changes that involve concerted motions of a large number of atoms may require significant trial and error. 'Undoing' the effect of a poorly chosen bias in an iMD-VR simulation often requires restarting the simulation and making another attempt.

Here we address these issues with a framework that combines interactive manipulation of ensembles of MD simulations in virtual reality (eMD-VR) with iterative batch jobs processing. We show successful sampling of a key intermediate along the folding pathway of a small protein in virtual reality by starting from an extended structure. We then show that eMD-VR generates plausible description of large scale loop motions in the enzyme cyclophilin A, and that the eMD-VR generated

pathways are suitable input to Markov State model workflows that allow rigorous calculation of thermodynamics and kinetics associated with a conformational change. This work opens up a new direction for interactive molecular dynamics simulations in situations where it is difficult or impractical to sample plausible pathways for conformational changes in real time. The focus of this work is to demonstrate proof of concept with a prototype implementation. We have not sought to quantify the extent to which VR visualization is better suited than alternative user interfaces, or whether the biasing protocols used here are optimal. Neither have we systematically benchmarked our approach against a diverse range of established enhanced sampling molecular dynamics methodologies or sought to assess the impact of prior user knowledge about the proteins studied on on the quality of the results obtained. Answers to such questions can be thoroughly researched once a robust prototype has become available.

**Implementation**

*User interface design*

We aimed to implement a goal-oriented sampling method where series of short (minutes to hours) MD simulations are carried out from starting coordinates selected by a user. This is a similar philosophy to families of adaptive sampling algorithms such as FAST where starting points for successive swarms of MD simulations are biased towards achieving a desired physical property (for instance minimizing radius of gyration as a proxy for encouraging folding of a protein structure).[70] The chief difference is that we rely on human visualization of MD trajectories and human intuition to guide the exploration of conformational space.

To this end we devised a user interface that would facilitate the management of a large number of MD simulations iteratively performed in an effort to discover a pathway for a desired conformational change. We also sought to exploit the power of VR visualization

to facilitate the visualization and analysis of processed MD trajectories, and the interactive setup and dispatch of MD simulations. We denote this approach as 'ensemble molecular dynamics virtual reality visualization' (eMD-VR) as it is different from the real-time single simulation centric approach which has been emphasized by iMD-VR methods.

To address these requirements the eMD-VR prototype was designed with a dual interface solution (Figure 1 and SI movie 1). In the trajectory manager layout the user initiates a new project by importing a structure or a previously computed MD trajectory (Figure 1a). In the current implementation, a tree-like structure is used to represent the relationship between the different MD simulations carried out over the course of the project. Each node (or leaf of the tree) corresponds to a single MD trajectory. Edges between two layers of nodes indicate that the MD trajectory of the lower layer node was initiated from a snapshot sampled from the upper layer MD trajectory. Therefore, this representation allows keeping track of the simulations hierarchy and to selectively collapse or expand specific branches corresponding to related simulations. A button menu enables the user to extend in one-click the sampling time of an individual MD trajectory, check the status of the submitted trajectories, or prepare the raw trajectory for visualization by removing periodic boundary artefacts and aligning snapshots on a common frame of a reference. The interface also allows deleting trajectories from the project. Different icons are used to represent the status of a node which can be for instance 'running' if a MD trajectory is being processed, or 'queued' if the trajectory has been setup but sampling hasn't yet started. To allow for transferability between the eMD-VR and other analysis software packages we implemented a feature to highlight visualized snapshots of particular relevance and export them as protein data bank (PDB) files.

Visualization of completed MD trajectories and spawning of new simulations is handled with a virtual reality graphical user interface (Figure 1b) that is accessed by clicking on one of the nodes in the trajectory manager interface. The current implementation makes use of an HTC vive headset and a dual controller system. The primary controller features a menu panel and the second controller is used as a pointer to execute different actions (Figure 1b). This dual controller approach provides a more intuitive way of rotating and positioning the MD trajectory than typical mouse-based visualizers by mimicking the way humans use both hands to manipulate objects. The user can scroll through the trajectory interactively and, by pointing and clicking with the secondary controller on a pictogram-based menu, can trigger different actions on the simulation. The actions available to the user in the current version of the prototype are: changing between a single atom selector and a complete residue selector, the possibility to switch between a ribbon representation and a combination of ribbon and lines and the ability to flag snapshots that are of particular interest. Importantly, the prototype allows selecting single structures (or snapshots) as a starting point for a new equilibrium or biased MD simulation (Figure 1c) and to spawn new simulations that are automatically incorporated as a new node in the trajectory manager layout. Steered molecular dynamics (SMD) is used to implement biasing potentials in eMD-VR (*vide infra*). To set up a biased simulation, the user simply selects by clicking with the controllers the residues that will feel the biasing potential and points and clicks to where the atoms will be pulled. By scrolling on the secondary controller the user can select the magnitude of the spring force constant required in SMD. The full capabilities of the prototype are detailed in SI movie 1.
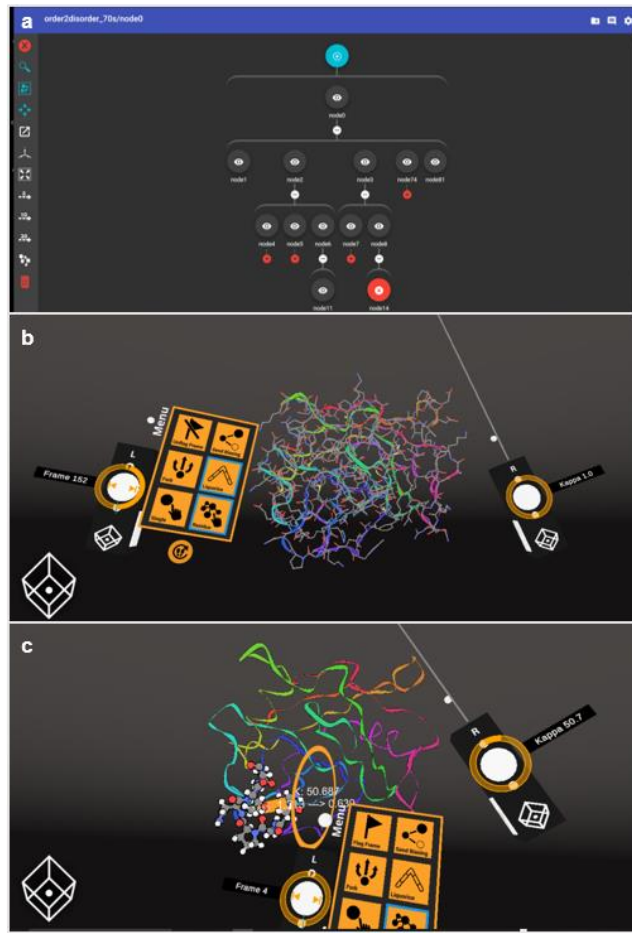
**Figure 1: User interface for eMD-VR projects. a** The web-browser trajectory manager interface is used to monitor, manage and process collections of MD simulations of a given system. Each simulation is depicted as a node in the tree and pictograms can be used to represent different states (such as ready to visualize, running, failed). Buttons on the left hand side are used as click triggers for routine post processing tasks (such as PBC artifact removal, alignment to a reference structure). **b** In-world VR GUI visualization of one completed MD trajectory using a dual controller virtual reality interface. Scrolling on the primary controller allows visualizing the trajectory forward and backwards. An icon based menu on the primary controller can be used to execute different actions as explained in the main text. **c** In-world VR GUI visualization depicting the setup of a biased MD simulation by selecting residues to be biased (that are highlighted by the prototype in licorice representation) to move towards a target region (orange circle) as explained in the main text.

*Software architecture*

eMD-VR was devised as three connected layers, which combine to enable the management and setup of an ensemble of biased MD trajectories (Figure 2). The top layer consists of two user interfaces (the trajectory manager and the VRGUI) developed with Unity software. Communication between the two interfaces is managed on the second layer with a Web API. In addition to handling the communication between the two UIs, the API connects them with a python based back-end that translates the commands issued on the user interfaces into GROMACS-PLUMED2[71,72] input files. The python back-end also handles remote deployment of molecular simulations on compute resources by using the Longbow library.[73] Both the API and the python back-end are stored in a Docker container.[74] This design was adopted to grant flexibility to execute MD simulations via eMD-VR on local or remote computing infrastructures.
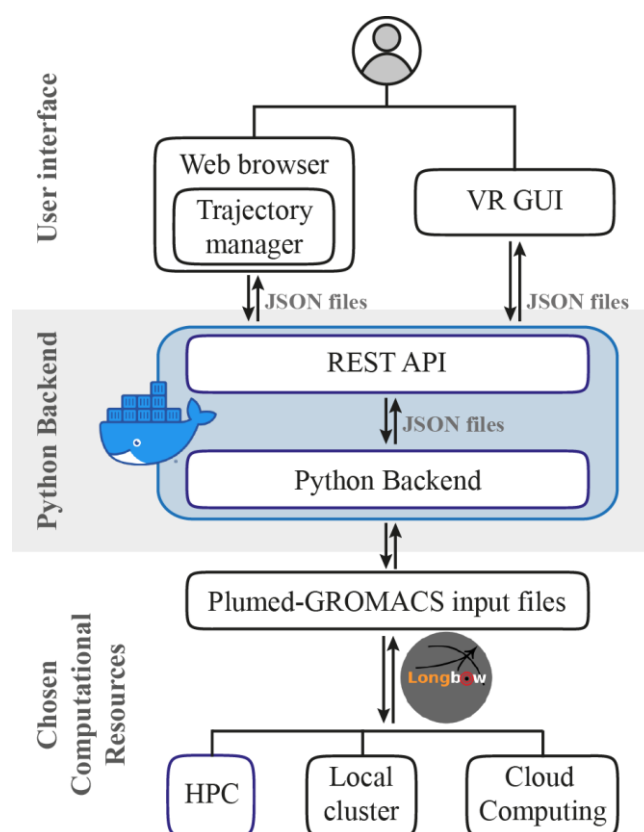


**Figure 2: Schematic representation of the three layers architecture of the iEBDD prototype.** See the main text for a detailed description of each layer.

*Implementation of eMD-VR simulations*

In eMD-VR, the implementation of biasing potentials is accomplished using SMD.[75] Specifically, the position of the center of mass (COM) of a set of atoms selected by the user is biased to adopt the position of a user-chosen virtual site whose coordinates are defined in a local frame of reference with respect to the protein. The present implementation uses a Harmonic Potential centered on a point which moves linearly with time (equation 1):

$$H(\boldsymbol{x}, t) = H(\boldsymbol{x}) + \kappa(s(\boldsymbol{x}, t) - s_0 - vt)^2 \ (1)$$

Where $H$ is the Hamiltonian of the system, $\boldsymbol{x}$ the position vector, $s(\boldsymbol{x},t)$ the center of mass of the selected atoms at time $t$, $s_0$ the center of mass at time zero, $v$ the pulling speed. In the eMD-VR prototype the value of $\kappa$ can be easily adjusted from the VR GUI using the scroll button on the secondary controller. The speed of pulling $v$ is determined from the displacement vector between $s_0$ and the position of the target COM, and the duration of the biased simulation, such that the center of mass of the biased atoms reaches the target position in the last 10% of the duration of the biased simulation.

All eMD-VR biased simulations were run in the NPT ensemble with a 2 fs timestep, using the leap-frog integrator and LINCS[76] algorithm to constrain bonds involving hydrogen, truncating the constraint coupling matrix at the fourth order. The generation of non-bonded pair lists was achieved with the Verlet scheme using a radius of 10 Å. Long range electrostatic interactions were handled using PME with a radius of 10 Å and a grid spacing of 1.6 Å, and Van der Waals interactions were handled using Lennard-Jones with a cut-off of 10 Å. Temperature was maintained using stochastic velocity rescaling coupling with a $\tau_t$ of 0.1 ps, and the Parrinello-Rahman barostat with a $\tau_p$ of 2 ps was used to maintain 1.0 bar constant pressure, with a system compressibility of $4.5e^{-5}$ bar$^{-1}$. A long range correction term was applied to correct for

the use of Van der Waals cutoff. Default values were used for the remaining parameters. MD simulations were set up and conducted using the appropriate modules of the GROMACS5.0[71] software package and PLUMED2[72] was used to incorporate the biasing potentials according to the parameters set up in the VR GUI.

*General procedure for generating conformational changes using eMD-VR*

Simulations in eMD-VR are typically started during VR visualization of a protein structure or a MD trajectory loaded in the software. If a PDB structure was provided as input, the software will automatically set up a basic simulation protocol consisting of minimisation, heating and density equilibration and a short simulation of 10ns. Once this simulation is finished, it can be loaded in the VR environment and visually inspected. From this simulation, the user can spawn as many simulations as needed from different points along the trajectory (both biased and non-biased simulations). Longbow will deploy these simulations to a remote computing node equipped with a SLURM workload manager. The software monitors automatically the progress of the remotely deployed simulations; hence the number of nodes that can be spawned simultaneously is only limited by the computational resources available to the user. Once the simulations are finished the output is retrieved by the software, processed to remove PBC artefacts and to align the snapshots to a reference structure, allowing the user to focus on visual analyses. The user can identify molecular structures of interest and highlight them using the flagging mechanic implemented with the controllers. These flagged structures can be exported and used in subsequent studies. For example, in the second test case of the current manuscript this feature was used to seed a swarm of MD trajectories (*vide infra*). The simulations files are accessible to the user and kept in their original GROMACS format, and can be readily analysed outside the prototype

using other software. Hence, the prototype considerably facilitates the procedure of setting up and performing MD simulations.

**Other Simulation Methodologies**

For the GTT protein we also carried out control non interactive equilibrium MD simulations using the same conditions as the default eMD-VR protocol, but with a sampling time of 1.5 μs. For studies of cyclophilin A (CypA) we carried out extensive follow-up MD simulations on eMD-VR generated pathways for motions of the 70s and 100s loops. Following visual inspection, 726 structures flagged along different paths generated with the VRGUI were used as initial structures to seed sets of MD trajectories. We attempted to evenly sample along trajectories were significant movements (*i.e* visually obvious) of either the 100s or the 70s loops were observed. The AMBER FF14SB forcefield was used and each protein structure was then solvated in a 12 Å of radius dodecahedral water box and 1 $Cl^-$ anion was added to neutralize the system, yielding 726 input conformations of approximately 30000 atoms. We followed the same minimization and equilibration protocols previously reported.[77] In brief, each system was energy minimized combining steepest descent and conjugate gradient (4500 and 500 iterations respectively) minimisation. Systems were then heated from 0 K to 250 K in 150 ps in the NVT ensemble using a time step of 0.5 fs, and from 250 to 298 K in 300 ps using a timestep of 1 fs. Before starting the production runs, each system was equilibrated for 300 ps at 298 K and 1 bar of pressure using a 2 fs timestep. Each system was then used for production runs consisting of 150 ns long trajectories in the NTP ensemble. All the MD simulations were set up and conducted using the appropriate the CUDA enabled version of PMEMD[78] and SHAKE[79] was applied to all bonds involving hydrogen. The AMBER hydrogen mass repartition scheme was applied to allow for a time step of 4 fs during the production MD simulations.[80]

The resulting pool of trajectories was used to construct a Markov State Model (MSM) using the pyEMMA 2.3.0 software package,[81] following our previously described protocol to build MSMs for this system.[77] In brief, the RMSD of the 70s loop with respect to the X-ray structure 1AK4[82] (residues Gly65 to Gly77) was used to monitor the movement of this loop. The movement of the 100s loop (residues Met100 to Ser110) was described monitoring simultaneously the RMSD of the loop, and the distance between the COM of the loop and the COM of the α-helix defined by residues Pro30 to Thr41. The 100s loop undergoes a flapping movement that maps upward and downward movements of the loop onto similar RMSD values. Following previous work,[77] we multiply RSMD values by -1 when the distance described above is shorter than that observed in the X-ray structure (14.0 Å) to separate both loop motions along this collective variable.

To allow the most direct comparison with our previously reported model, we also used the same clustering algorithm (K-means) and the same number of microstates (100) from our previous work[77] for the MSM construction. Different lagtimes were used to compute implied timescales of the dominant eigenvectors, and the resulting implied timescale plots are shown in Figure S1. A Bayesian MSM was built at each lagtime in order to obtain an error estimate for the obtained timescales. In agreement with our previous study a lagtime of 40 ns was chosen for the MSM construction, with default parameters provided by PyEMMA. Coarse-graining of the resulting MSM was also made using the same approach as reported in our previous work: the 100 k-means derived model was coarse grained into 5 metastable states according to the following structural criteria: microstates with a value of the 70s-loop CV below 1.5 Å and with a value of the 100s-loop CV below 4.5 Å were assigned to the ground state (closed/open, orange). Microstates within the same cutoff of the 70s-loop CV but with values of the

100s-loop above 4.5 Å were assigned to the closed/closed metastable state (red). Microstates with values of the 70s-loop CV between 1.5 Å and below 4.0 Å were assigned to the intermediate macro-state (teal). Microstates with a 70s-loop CV value above 4.0 Å and a 100s-loop CV value < 4.5 Å were assigned to the open/open state (magenta) while microstates with similar values of the 70s-loop CV but extreme values of the 100s-loop CV (>4.5 Å) were assigned to the open/closed (blue) macro-state.

**Computational resources.**

The VR-environment was hosted on a Dell Alienware 15R3 laptop equipped with a 2.8 GHz Intel i7 core16 Gb of RAM and an Nvidia GeForce GTX 1070 graphics processing unit. The remote computing node was an Armari Magnetar workstation equipped with 20 Intel Xeon processors, 32 Gb of RAM and 4 Nvidia GeForce GTX 980 Ti graphic processing units. Production runs for the swarm of trajectories were performed combining a local cluster and the Archer and Cirrus UKs National HPC facilities.

**Results and Discussion**

**eMD-VR accelerates protein folding simulations**

The GTT protein is an engineered variant of the WW domain FiP35 (sequence: GSKLPPGWEKRMSRDGRVYYFNHITGTTQFERPSG). MD simulations have demonstrated folding times in good agreement with estimates of 4 μs derived from T-jump experiments carried out at a temperature of 353 K, close to the melting point of GTT.[83] At a room temperature of 298 K the folding time is anticipated to be closer to the millisecond timescale.[84] Indeed Nguyen et. al.[85] could only fold this sequence starting from a fully extended structure at 325 K by using an implicit solvent model to reduce friction.

To test the potential applicability of eMD-VR to quickly explore protein folding pathways, we aimed to explore the folding of GTT in explicit solvent at room temperature starting from a fully extended structure (Figure 3). With typical settings, each iteration of the workflow had a wallclock time of approximately 20 minutes (details of the settings used at each iteration are summarised in Table S1). We focused on the formation of hairpin 1 (residues 8 to 23) as formation of this hairpin has been suggested to be the dominant intermediate in the folding pathway.[86] According to the crystal structure of this WW domain (PDB ID 2F21),[87] the hairpin 1 β-sheet structure is characterized by a series of close contacts pairs between the alpha carbons of the antiparallel strands. Specifically, the following Cα-Cα distances were measured in the crystallographic structure for residue pairs 8-22 (4.2 Å), 9-21 (5.4 Å), 10-20 (4.7 Å), 11-19 (5.4 Å), 12-18 (4.5 Å), 13-17 (5.3 Å) and 14-16 (5.7 Å). To reproduce this pattern of contacts, we first attempted to bring the N and C terminal regions of the hairpin closer to each other. However, after just one iteration (amounting to 10 ns of sampling time), the extended structure underwent hydrophobic collapse and the RMSD rapidly decreased from initial values of ca. 9.8 Å down to ca 4.8 Å. However, the root mean square deviation for the reference contact pairs ($RMSD_{C\alpha-C\alpha}$) was 6.0 Å, consistent with a misfolded structure of the hairpin 1 strands. Therefore, a bias was applied to unwind the structure, and the structure was refolded whilst keeping the correct alignment of the antiparallel strands. Following this protocol we reached intermediate II, which displayed a partially folded structure (RMSD 4.0 Å, $RMSD_{C\alpha-C\alpha}$ 3.3 Å). From this structure, three different pathways were attempted to form hairpin 1, by forming the pairs 13-17, 12-18, and 10-20, leading to structures III, IV and V with RMSD values of 4.2 / 5.3 / 3.5 Å and $RMSD_{C\alpha-C\alpha}$ values of 3.1 Å, 6.1 Å, and 2.0 Å, respectively. Refinement of these structures was sought using eMD-VR steps of unbiased MD

simulations, leading to structure VI with RMSD of 2.0 Å and $RMSD_{C\alpha-C\alpha}$ of 0.7 Å. This level of agreement for structure VI is comparable to results previously reported in the literature for both multi μs[85] and enhanced sampling simulations.[88] However the overall sampling time for the successful pathway was only 64.8 ns. By contrast a control equilibrium MD simulation of 1.5 μs could only achieve values of RMSD 6.8 Å and $RMSD_{C\alpha-C\alpha}$ 9.9 Å (Table S2).
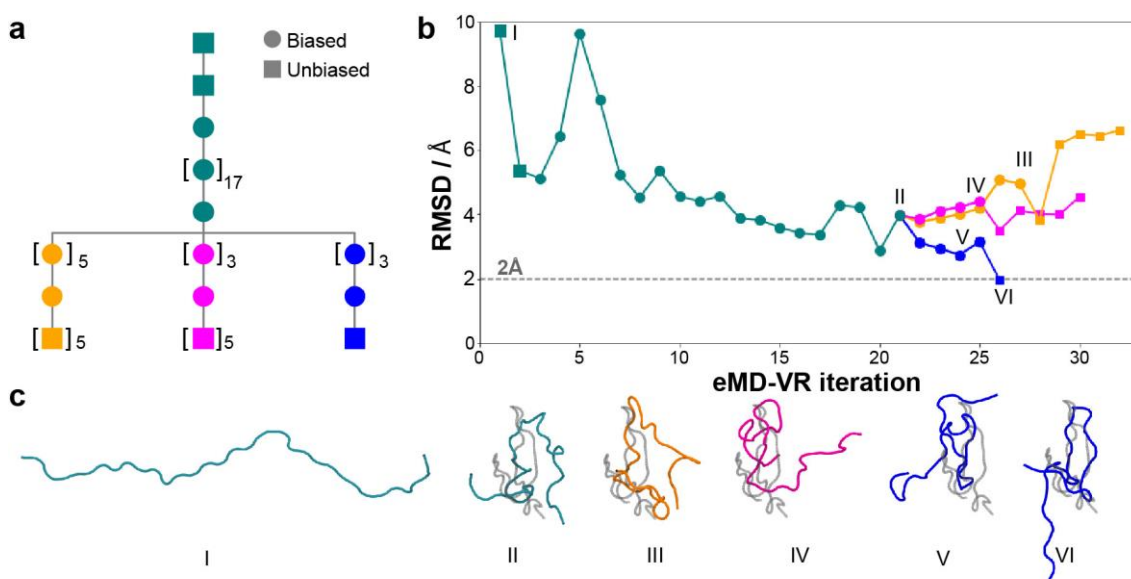


**Figure 3: Folding pathway of the GTT protein with eMD-VR. a** Schematic view of the different pathways explored with the eMD-VR interface. Numbers in bold represent the number of eMD-VR iterations. **b** Root Mean Squared Deviation with respect to the fully folded hairpin 1 of the WW domain in PDB 2F21 (residues 8 to 23) along multiple iterations of eMD-VR. **c** Representative structures of conformational states sampled along the eMD-VR trajectories.

**eMD-VR generates realistic pathways for millisecond timescale loop motions**

We have recently reported an all-atom computational model of the conformational change that occurs between the native state and a transient state of the protein Cyclophilin A (CypA), a key enzyme for the life cycle of the HIV virus.[77] This motion

involves large scale rearrangements of a so called 70s loop region from a closed to an open state, and a so called 100s loop region from an open to a closed state (Figure 4a). The loop motions occur on micro to millisecond timescale and are critically dependent on interactions that Asp66 establishes with the 70s loop residues, as was demonstrated by NMR experiments on wild-type and mutant forms of CypA. The computational model was generated by combining the accelerated MD method (aMD) to sample plausible loop motion pathways, with Markov State Modelling (MSM) to obtain information about the thermodynamics and kinetics of the process for a more quantitative comparison with experimental data.

A limitation of the aMD methodology is that the acceleration of rare events depends on the selection of a so called ''boost'' parameter that can be cumbersome to optimise. Too little boost fails to significantly enhance sampling of conformations over the timescale of a ~few hundred ns long MD simulation; whereas too much boost causes the protein to unfold. In previous work we carefully optimised the boost level via trial and errors over a period of a few weeks to observe enhanced motions of the 70s and 100s loop regions, requiring microseconds of aggregate aMD sampling time, whilst adding positional restraints to key regions of CypA to prevent global unfolding.

We sought to explore whether eMD-VR could be a more efficient alternative to aMD to generate large scale movements of the flexible loops of CypA. Starting from the crystal structure of CypA (PDB ID 1AK4) we followed a stepwise approach to trigger the conformational changes in both loops. A total of 26 and 17 iterations respectively were required to observe a movement of the 70s loop from a closed to an open state, and movements of the 100s loop from an open to a closed state, representing a total MD sampling time of less than100 ns.

Encouragingly inspection of the eMD-VR generated pathways indicates that the opening of the 70s loop occurs via a similar pathway to that observed in the aMD simulations (Figure 4b). In particular both pathways progress via a partially unwound 70s loop structure where approximately half of the hydrogen bonding interactions between Asp66 and neighbouring 70s loop residues are broken, before progressing to a fully open loop conformation that no longer interacts with Asp66. Moreover, with both aMD and eMD-VR protocols the changes in number of hydrogen bonds between open and closed states in the vicinity of the 70s loop region indicate the important role of Asp66 (Figure 4c). This suggests that eMD-VR generated pathways are suitable to generate molecular design hypotheses. However, comparison of hydrogen bonding patterns along aMD and eMD-VR pathways revealed the breaking of the interactions of Asp66 with the 70s loop residues is more abrupt following the eMD-VR protocol (Figure S3).
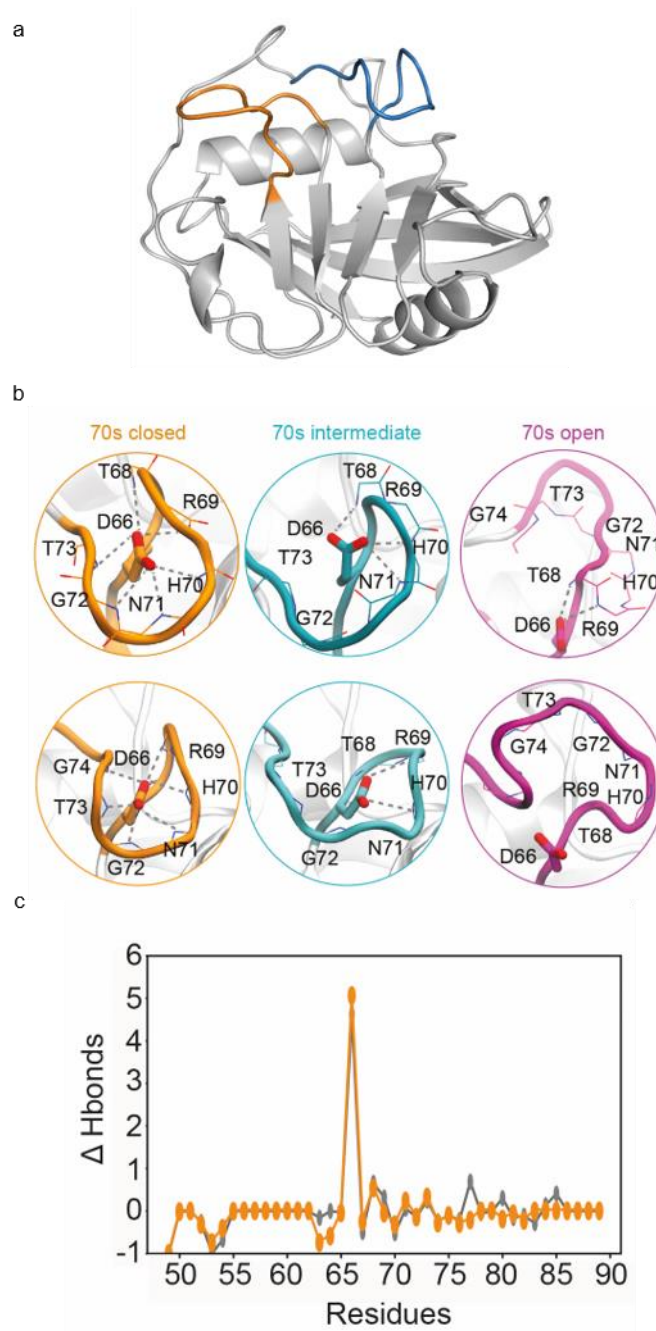
**Figure 4: Exploring large amplitude loop movements in CypA a.** Location and conformation of the 70s (blue) and 100s (orange) loops in the crystallographic structure of CypA (PDB: 1AK4). **b.** Metastables states of the 70s loop identified using the aMD (top) and eMD-VR (bottom) protocols. **c.** The difference in the average number of intra-molecular H-bonds in the 70s closed and 70s open states as previously identified using aMD (grey) and as identified using the eMD-VR protocol (orange). In both cases residue D66 stands out as it has substantially more H-bonds stabilising the closed state than the open state of the 70s loop.

Thus to further assess the validity of the eMD-VR generated pathways we used the structures sampled along the pathways describing opening of the 70s or the 100s loop to seed a swarm of equilibrium MD simulations. Additionally, three simulations started from the initial structure and each of the two end-points of the sampled pathways, were simulated for 2 μs each. The resulting sets of simulations were then processed using our Markov state model (MSM) analysis workflow described elsewhere.[77] Briefly, the set of MD trajectories was coarse-grained on a conformational ensemble using five representative macrostates. These five macrostates were in good agreement with the conformational ensemble we had previously described. The superimposition of the average structures obtained for these macrostates and the $C_\alpha$ RMSD values with respect to the average structure of the corresponding state obtained using the aMD based protocol are displayed in Figure 5a. Furthermore, the distributions of RMSD values of the 70s and 100s loops with respect to the CypA crystal structure was also very similar among the different macrostates, even though the amplitude of the 100s loop distribution was slightly narrower in the results from the eMD-VR protocol (Figure 5b and 5c). These similarities suggest that the conformational plasticity within these macrostates is qualitatively the same independently of whether the initial seeding was conducted with aMD or eMD-VR. Finally, we compared the populations of macrostates and estimated mean first passage time values (MFPT) for the transitions between macrostates (Figure 6) in the two models. Both models indicate that the 70s loop adopts a predominantly closed state, but with a slightly greater preponderance of 70s open states in the eMD-VR/MSM model with respect to the original model (ca. 70%/25% vs 60%/35%). These differences in macrostates population are however within the uncertainties of the respective MSMs. Both MSMs identify the key low-population intermediate structure (cyan) for rearrangements of the 70s loop. In both cases the

conformational rearrangements of the 100s loop were predicted to occur within a 1 to 10 μs timescale. The opening of the 70s loop was predicted to occur between 3 and 6 times faster using the eMD-VR seeded pathways. This level of agreement between the two MSMs was deemed satisfactory given the overall uncertainties on the computed mean first passage times. The eMD-VR/MSM protocol used approximately half the aggregated sampling time of the aMD/MSM protocol, but it is apparent that this was sufficient to reach the same qualitative conclusions about the relative importance of the distinct conformational states of the 70s and 100s loops, and the nature of the slowest process. Further agreement could be sought via extending the simulation time.
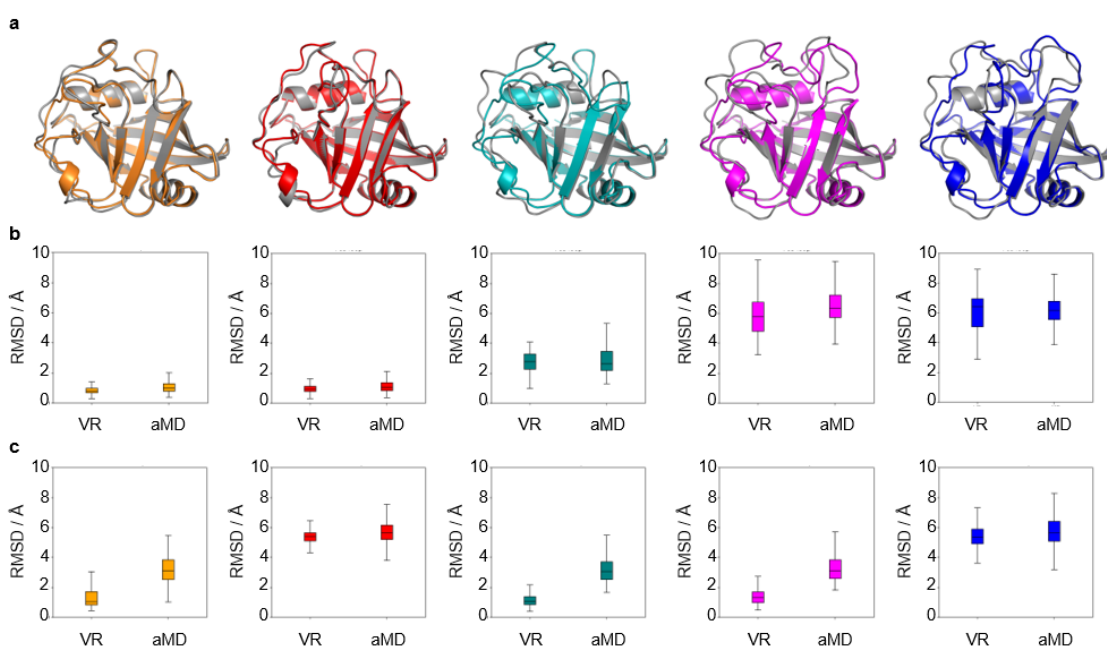


**Figure 5: a.** Superposition of the average structure of each macrostate with the corresponding state described elsewhere.[77] **b.** Per macrostate distribution of RMSD values of the 70s loop with respect the crystallographic structure with PDB ID 1AK4 for structures sampled with the VR, or aMD protocols. **c.** Per macrostate distribution of RMSD values of the 100s loop with respect the crystallographic structure with PDB ID 1AK4 for structures sampled with the VR, or aMD protocols.
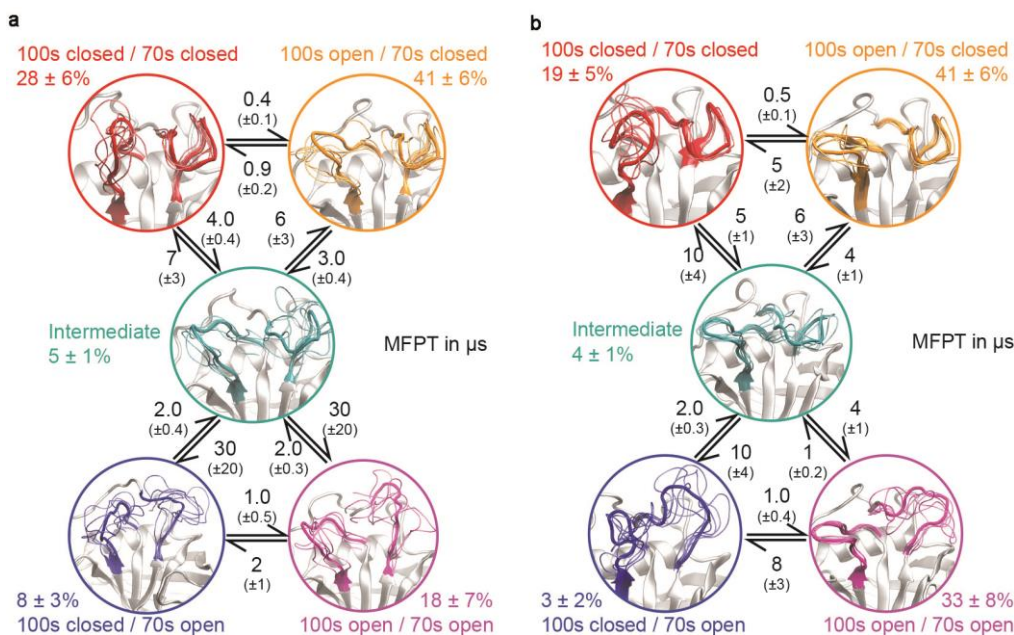
**Figure 6**. Models of CypA loop dynamics. **a.** Obtained with the aMD/MSM protocol. **b.** Obtained with the eMD-VR/MSM protocol. Error bars on reported populations and mean first passage times (MFPT) were obtained by bootstrapping of the MD trajectories assigned to the individual microstates.

**Conclusion**

We have described a new VR-based method to conduct MD simulations, incorporating interactive biases to explore large amplitude conformational changes in protein structures. Recent results have shown the potential of iMD-VR simulations to study a range of different molecular systems.[38,59] In this work, we have extended those ideas to explore ensemble-based VR-enabled visualization strategies as a strategy for launching steered molecular dynamics simulations. By introducing the trajectory manager interface, we enabled simultaneous exploration of multiple pathways, and facilitated back-tracking of interactively biased simulations. The design of the VR GUI as a setup/submit/recover simulation protocol allows quick access to a range of simulation data, which can be obtained either through conventional simulation approaches, interactively biased MD, or steered MD. Steered MD in particular allows for the use of gentle biases to trigger conformational changes, which may enable the exploration of

pathways that remain close to minimum energy pathway. We have demonstrated that with this implementation, eMD-VR is a suitable alternative to aMD simulations to explore the loop dynamics of CypA, achieving comparable conformational changes with ca. one order of magnitude less sampling. A powerful feature of the prototype is that it allows non-expert users to intuitively setup, execute and visualise a large number of MD simulations. We have received positive feedback about the potential of eMD-VR via running short research projects for visiting high school students, and via live demos at the 2019 Edinburgh Science Festival and Open Days at the University of Edinburgh.

eMD-VR offers potential for exploring protein folding pathways and obtain insights into the nature of soft condensed matter at the nanoscale. However we stress that, given the subjective nature of the visual analyses, there may be variability in the results obtained with the prototype by different users. The hypotheses generated with the aid of the software should be assessed with complementary approaches (such as Markov state modeling or other energetic analyses). Further development could improve the efficiency of eMD-VR studies as a number of trial and error attempts were needed to fold the first hairpin of the GTT protein. In future versions of the prototype we aim to implement the setup of positional restraints from within the VR GUI, as it could ease the folding of more complex secondary structure elements, and reduce the number of interactions the user has to undertake to achieve a desired conformational change. We have also observed that the opening of the 70s loop happened more abruptly in the eMD-VR case than in the aMD case, suggesting that more gentle biasing potentials could improve further the agreement between the methodologies. We note that the current prototype could be also deployed in a non-VR environment, with interactive visualization and setup of biased MD simulations done via a touch-screen or monitor and keyboard interface. Previous studies have suggested that VR-based

environments accelerate substantially the execution of a variety of molecular modelling tasks,[38] and in future work we plan to carry out an in-depth human-computer interaction study to quantify the effectiveness of eMD-VR over other interfaces for ensemble MD studies. It would also be beneficial to implement a MD-package agnostic back-end via a framework such as BioSimSpace to facilitate deployment of the prototype on a range of diverse compute resources.[89] Potential scientific developments of interest for the prototype include the implementation of distance measurement and energetic estimation that could help improve the reliability of the results obtained. Further, integration of eMD-VR with druggability predictors such as JEDI,[90] or Fpocket,[91,92] to allow the interactive evaluation of protein druggability, which could facilitate the sampling of cryptic pockets for drug discovery purposes.[93] Overall the present results indicate the potential of eMD-VR to broaden access to simulations of complex conformational changes in proteins.

**Software availability**

Input structures employed for the GTT and CypA test cases are available for download at https://github.com/michellab/eMD-VR-inputs. An evaluation version of the software is available upon request.

**Acknowledgements**

**Conflict of Interest**

PT and RS are employees of the company Interactive Scientific that commercializes virtual-reality visualization software. JM is a current member of the Scientific Advisory Board of Cresset. These materials are not sponsored by or affiliated with Unity Technologies or its affiliates. "Unity" is a trademark or registered trademark of Unity Technologies or its affiliates in the U.S. and elsewhere.

**References**

(1)     Karplus, M. Molecular Dynamics Simulations of Biomolecules. *Accounts of Chemical Research* **2002**, *35*, 321-323.

(2)     Pérez, A.; Luque, F. J.; Orozco, M. Frontiers in Molecular Dynamics Simulations of DNA. *Accounts of Chemical Research* **2012**, *45*, 196-205.

(3)     Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* **2012**, *41*, 429-452.

(4)     Cheatham III, T. E.; Case, D. A. Twenty-five years of nucleic acid simulations. *Biopolymers* **2013**, *99*, 969-977.

(5)     Maximova, T.; Moffatt, R.; Ma, B.; Nussinov, R.; Shehu, A. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLoS Computational Biology* **2016**, *12*, e1004619.

(6)     De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *Journal of medicinal chemistry* **2016**, *59*, 4035-4061.

(7)     Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein–RNA complexes? *WIREs RNA* **2017**, *8*, e1405.

(8)     Bottaro, S.; Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* **2018**, *361*, 355-360.

(9)     Gundogdu, M.; Llabrés, S.; Gorelik, A.; Ferenbach, A. T.; Zachariae, U.; van Aalten, D. M. F. The O-GlcNAc Transferase Intellectual Disability Mutation L254F Distorts the TPR Helix. *Cell Chemical Biology* **2018**, *25*, 513-518.e514.

(10)    Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W. Biomolecular

simulations: From dynamics and mechanisms to computational assays of biological activity. *WIREs Computational Molecular Science* **2019**, *9*, e1393.

(11)    Palermo, G.; Casalino, L.; Magistrato, A.; Andrew McCammon, J. Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *Journal of Structural Biology* **2019**, *206*, 267-279.

(12)    Mereghetti, P.; Kokh, D.; McCammon, J. A.; Wade, R. C. Diffusion and association processes in biological systems: theory, computation and experiment. *BMC Biophys* **2011**, *4*, 2.

(13)    Meng, Y.; Gao, C.; Clawson, D. K.; Atwell, S.; Russell, M.; Vieth, M.; Roux, B. Predicting the Conformational Variability of Abl Tyrosine Kinase using Molecular Dynamics Simulations and Markov State Models. *Journal of Chemical Theory and Computation* **2018**, *14*, 2721-2732.

(14)    Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964-972.

(15)    Wang, Y.; Papaleo, E.; Lindorff-Larsen, K. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *eLife* **2016**, *5*, e17505.

(16)    Lee, C. T.; Amaro, R. E. Exascale Computing: A New Dawn for Computational Biology. *Computing in Science & Engineering* **2018**, *20*, 18-25.

(17)    Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C.: Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE Press: New Orleans, Louisana, 2014; pp 41-53.

(18)    Kästner, J. Umbrella sampling. *WIREs Computational Molecular Science* **2011**, *1*, 932-942.

(19)    Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc Natl Acad Sci U S A* **2002**, *99*, 12562-12566.

(20)    St-Pierre, J.-F.; Karttunen, M.; Mousseau, N.; Róg, T.; Bunker, A. Use of Umbrella Sampling to Calculate the Entrance/Exit Pathway for Z-Pro-Prolinal Inhibitor in Prolyl Oligopeptidase. *Journal of Chemical Theory and Computation* **2011**, *7*, 1583-1594.

(21)    Kokubo, H.; Tanaka, T.; Okamoto, Y. Ab Initio prediction of protein–ligand binding structures by replica-exchange umbrella sampling simulations. *Journal of Computational Chemistry* **2011**, *32*, 2810-2821.

(22)    Zhang, Y.; Voth, G. A. Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles. *Journal of Chemical Theory and Computation* **2011**, *7*, 2277-2283.

(23)     Granata, D.; Camilloni, C.; Vendruscolo, M.; Laio, A. Characterization of the free-energy landscapes of proteins by NMR-guided metadynamics. *Proceedings of the National Academy of Sciences* **2013**, *110*, 6817-6822.

(24)     Deighan, M.; Pfaendtner, J. Exhaustively Sampling Peptide Adsorption with Metadynamics. *Langmuir* **2013**, *29*, 7999-8009.

(25)     Kokubo, H.; Tanaka, T.; Okamoto, Y. Prediction of Protein–Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations: Application to Kinase Systems. *Journal of Chemical Theory and Computation* **2013**, *9*, 4660-4671.

(26)     Wang, J.; Shao, Q.; Xu, Z.; Liu, Y.; Yang, Z.; Cossins, B. P.; Jiang, H.; Chen, K.; Shi, J.; Zhu, W. Exploring Transition Pathway and Free-Energy Profile of Large-Scale Protein Conformational Change by Combining Normal Mode Analysis and Umbrella Sampling Molecular Dynamics. *The Journal of Physical Chemistry B* **2014**, *118*, 134-143.

(27)     Bochicchio, D.; Panizon, E.; Ferrando, R.; Monticelli, L.; Rossi, G. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *The Journal of Chemical Physics* **2015**, *143*, 144108.

(28)     Bueren-Calabuig, J. A.; Michel, J. Elucidation of ligand-dependent modulation of disorder-order transitions in the oncoprotein MDM2. *PLoS Computational Biology* **2015**, *11*, e1004282.

(29)     Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug–Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Accounts of Chemical Research* **2015**, *48*, 277-285.

(30)     Bueren-Calabuig, J. A.; Michel, J. Impact of Ser17 Phosphorylation on the Conformational Dynamics of the Oncoprotein MDM2. *Biochemistry* **2016**, *55*, 2500-2509.

(31)     Llabrés, S.; Juárez-Jiménez, J.; Masetti, M.; Leiva, R.; Vázquez, S.; Gazzarrini, S.; Moroni, A.; Cavalli, A.; Luque, F. J. Mechanism of the Pseudoirreversible Binding of Amantadine to the M2 Proton Channel. *Journal of the American Chemical Society* **2016**, *138*, 15345-15358.

(32)     Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *Journal of the American Chemical Society* **2017**, *139*, 4780-4788.

(33)     Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T. An Efficient Metadynamics-Based Protocol To Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *Journal of Chemical Information and Modeling* **2017**, *57*, 1210-1217.

(34)     Kuzmanic, A.; Sutto, L.; Saladino, G.; Nebreda, A. R.; Gervasio, F. L.; Orozco, M. Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *eLife* **2017**, *6*, e22175.

(35)     Domański, J.; Hedger, G.; Best, R. B.; Stansfeld, P. J.; Sansom, M. S. P. Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association. *The Journal of Physical Chemistry B* **2017**, *121*, 3364-3375.

(36)     Meral, D.; Provasi, D.; Filizola, M. An efficient strategy to estimate thermodynamics and kinetics of G protein-coupled receptor activation using metadynamics and maximum caliber. *The Journal of chemical physics* **2018**, *149*, 224101-224101.

(37)     Wapeesittipan, P.; Mey, A. S. J. S.; Walkinshaw, M. D.; Michel, J. Allosteric effects in cyclophilin mutants may be explained by changes in nano-microsecond time scale motions. *Communications Chemistry* **2019**, *2*, 41.

(38)     O'Connor, M.; Deeks, H. M.; Dawn, E.; Metatla, O.; Roudaut, A.; Sutton, M.; Thomas, L. M.; Glowacki, B. R.; Sage, R.; Tew, P.; Wonnacott, M.; Bates, P.; Mulholland, A. J.; Glowacki, D. R. Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Science Advances* **2018**, *4*, eaat2731.

(39)     Dreher, M.; Piuzzi, M.; Turki, A.; Chavent, M.; Baaden, M.; Férey, N.; Limet, S.; Raffin, B.; Robert, S. Interactive Molecular Dynamics: Scaling up to Large Systems. *Procedia Computer Science* **2013**, *18*, 20-29.

(40)     Haag, M. P.; Vaucher, A. C.; Bosson, M.; Redon, S.; Reiher, M. Interactive Chemical Reactivity Exploration. *ChemPhysChem* **2014**, *15*, 3301-3319.

(41)     Luehr, N.; Jin, A. G. B.; Martínez, T. J. Ab Initio Interactive Molecular Dynamics on Graphical Processing Units (GPUs). *Journal of Chemical Theory and Computation* **2015**, *11*, 4536-4544.

(42)     Limniou, M.; Roberts, D.; Papadopoulos, N. Full immersive virtual environment CAVETM in chemistry education. *Computers & Education* **2008**, *51*, 584-593.

(43)     Cai, S.; Wang, X.; Chiang, F.-K. A case study of Augmented Reality simulation system application in a chemistry course. *Computers in Human Behavior* **2014**, *37*, 31-40.

(44)     Norrby, M.; Grebner, C.; Eriksson, J.; Boström, J. Molecular Rift: Virtual Reality for Drug Designers. *Journal of Chemical Information and Modeling* **2015**, *55*, 2475-2484.

(45)     Salvadori, A.; Del Frate, G.; Pagliai, M.; Mancini, G.; Barone, V. Immersive virtual reality in computational chemistry: Applications to the analysis of QM and MM data. *International Journal of Quantum Chemistry* **2016**, *116*, 1731-1746.

(46)     Zheng, M.; Waller, M. P. ChemPreview: an augmented reality-based molecular interface. *Journal of Molecular Graphics and Modelling* **2017**, *73*, 18-23.

(47)     Salvadori, A.; Fusè, M.; Mancini, G.; Rampino, S.; Barone, V. Diving into chemical bonding: An immersive analysis of the electron charge rearrangement through virtual reality. *Journal of Computational Chemistry* **2018**, *39*, 2607-2617.

(48)     Bennie, S. J.; Ranaghan, K. E.; Deeks, H.; Goldsmith, H. E.; O'Connor, M. B.; Mulholland, A. J.; Glowacki, D. R. Teaching Enzyme Catalysis Using Interactive Molecular Dynamics in Virtual Reality. *Journal of Chemical Education* **2019**, *96*, 2488-2496.

(49)     Amabilino, S.; Bratholm, L. A.; Bennie, S. J.; Vaucher, A. C.; Reiher, M.; Glowacki, D. R. Training Neural Nets To Learn Reactive Potential Energy Surfaces Using Interactive Quantum Chemistry in Virtual Reality. *The Journal of Physical Chemistry A* **2019**, *123*, 4486-4499.

(50)     Doak, D. G.; Denyer, G. S.; Gerrard, J. A.; Mackay, J. P.; Allison, J. R. Peppy: A virtual reality environment for exploring the principles of polypeptide structure. *Protein Science* **2020**, *29*, 157-168.

(51)     Probst, D.; Reymond, J. L. Exploring DrugBank in Virtual Reality Chemical Space. *J Chem Inf Model* **2018**, *58*, 1731-1735.

(52)     Laureanti, J.; Brandi, J.; Offor, E.; Engel, D.; Rallo, R.; Ginovska, B.; Martinez, X.; Baaden, M.; Baker, N. A. Visualizing biomolecular electrostatics in virtual reality with UnityMol-APBS. *Protein Sci* **2020**, *29*, 237-246.

(53)     Cassidy, K. C.; Šefčík, J.; Raghav, Y.; Chang, A.; Durrant, J. D. ProteinVR: Web-based molecular visualization in virtual reality. *PLoS Computational Biology* **2020**, *16*, e1007747.

(54)     Martinez, X.; Chavent, M.; Baaden, M. Visualizing protein structures - tools and trends. *Biochem Soc Trans* **2020**.

(55)     Ratamero, E. M.; Bellini, D.; Dowson, C. G.; Romer, R. A. Touching proteins with virtual bare hands : Visualizing protein-drug complexes and their dynamics in self-made virtual reality using gaming hardware. *J Comput Aided Mol Des* **2018**, *32*, 703-709.

(56)     O'Connor, M. B.; Bennie, S. J.; Deeks, H. M.; Jamieson-Binnie, A.; Jones, A. J.; Shannon, R. J.; Walters, R.; Mitchell, T. J.; Mulholland, A. J.; Glowacki, D. R. Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. *J Chem Phys* **2019**, *150*, 220901.

(57)     Ferina, J.; Daggett, V. Visualizing Protein Folding and Unfolding. *J Mol Biol* **2019**, *431*, 1540-1564.

(58)     Ventola, C. L. Virtual Reality in Pharmacy: Opportunities for Clinical, Research, and Educational Applications. *P T* **2019**, *44*, 267-276.

(59)     O'Connor, M. B.; Bennie, S. J.; Deeks, H. M.; Jamieson-Binnie, A.; Jones, A. J.; Shannon, R. J.; Walters, R.; Mitchell, T. J.; Mulholland, A. J.; Glowacki, D. R. Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. *The Journal of Chemical Physics* **2019**, *150*, 220901.

(60)     Deeks, H. M.; Walters, R. K.; Hare, S. R.; O'Connor, M. B.; Mulholland, A. J.; Glowacki, D. R. Interactive molecular dynamics in virtual reality for accurate flexible protein-ligand docking. *PloS One* **2020**, *15*, e0228461.

(61)     Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology* **2017**, *13*, e1005659.

(62)     Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **2013**, *3*, 198-210.

(63)     Xiong, H.; Crespo, A.; Marti, M.; Estrin, D.; Roitberg, A. E. Free Energy Calculations with Non-Equilibrium Methods: Applications of the Jarzynski Relationship. *Theoretical Chemistry Accounts* **2006**, *116*, 338-346.

(64)    Muller, M. P.; Jiang, T.; Sun, C.; Lihan, M.; Pant, S.; Mahinthichaichan, P.; Trifan, A.; Tajkhorshid, E. Characterization of Lipid–Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation. *Chemical reviews* **2019**, *119*, 6086-6161.

(65)    Hadden, J. A.; Perilla, J. R.; Schlicksup, C. J.; Venkatakrishnan, B.; Zlotnick, A.; Schulten, K. All-atom molecular dynamics of the HBV capsid reveals insights into biological function and cryo-EM resolution limits. *eLife* **2018**, *7*, e32478.

(66)    Amaro, R. E.; Mulholland, A. J. Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nature Reviews Chemistry* **2018**, *2*, 0148.

(67)    Khalid, S.; Rouse, S. L. Simulation of subcellular structures. *Current opinion in structural biology* **2020**, *61*, 167-172.

(68)    Reddy, T.; Shorthouse, D.; Parton, Daniel L.; Jefferys, E.; Fowler, Philip W.; Chavent, M.; Baaden, M.; Sansom, Mark S. P. Nothing to Sneeze At: A Dynamic and Integrative Computational Model of an Influenza A Virion. *Structure* **2015**, *23*, 584-597.

(69)    Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Ieong, P. U.; Dommer, A. C.; Amaro, R. E. Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent Sci* **2020**, *6*, 189-196.

(70)    Zimmerman, M. I.; Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *Journal of Chemical Theory and Computation* **2015**, *11*, 5747-5757.

(71)    Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.

(72)    Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014**, *185*, 604-613.

(73)    Gebbie-Rayet, J.; Shannon, G.; Loeffler, H. H.; Laughton, C. A. Longbow: A Lightweight Remote Job Submission Tool. *Journal of Open Research Software* **2016**, *4*, p.e1.

(74)    Merkel, D. Docker: lightweight linux containers for consistent development and deployment. *Linux Journal* **2014**.

(75)    Lu, H.; Schulten, K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins: Structure, Function, and Bioinformatics* **1999**, *35*, 453-463.

(76)    Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **2008**, *4*, 116-122.

(77)    Juárez-Jiménez, J.; Gupta, A.; Karunanithy, G.; Mey, A.; Georgiou, C.; Ioannidis, H.; De Simone, A.; Barlow, P.; Hulme, A. N.; Walkinshaw, M. D.; Baldwin, A. J.; Michel, J. Dynamic design: manipulation of millisecond timescale motions on the energy landscape of Cyclophilin A. *Chemical Science* **2020**.

(78)    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. *Journal of Chemical Theory and Computation* **2013**, *9*, 3878-3888.

(79)    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.

(80)    Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2015**, *11*, 1864-1874.

(81)    Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525-5542.

(82)    Gamble, T. R.; Vajdos, F. F.; Yoo, S.; Worthylake, D. K.; Houseweart, M.; Sundquist, W. I.; Hill, C. P. Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid. *Cell* **1996**, *87*, 1285-1294.

(83)    Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational Design and Experimental Testing of the Fastest-Folding β-Sheet Protein. *Journal of Molecular Biology* **2011**, *405*, 43-48.

(84)    Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. The folding mechanism of a b-sheet: the WW domain. *Journal of Molecular Biology* **2001**, *311*, 373-393.

(85)    Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *Journal of the American Chemical Society* **2014**, *136*, 13959-13962.

(86)    a Beccara, S.; Škrbić, T.; Covino, R.; Faccioli, P. Dominant folding pathways of a WW domain. *Proceedings of the National Academy of Sciences* **2012**, *109*, 2330-2335.

(87)    Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. Structure–function–folding relationship in a WW domain. *Proceedings of the National Academy of Sciences* **2006**, *103*, 10648-10653.

(88)    Shao, Q.; Shi, J.; Zhu, W. Determining Protein Folding Pathway and Associated Energetics through Partitioned Integrated-Tempering-Sampling Simulation. *Journal of Chemical Theory and Computation* **2017**, *13*, 1229-1243.

(89)    Hedges, L. O.; Mey, A. S. J. S.; Laughton, C. A.; Gervasio, F. L.; Mulholland, A. J.; Woods, C. J.; Michel, J. BioSimSpace: An interoperable Python framework for biomolecular simulation. *Journal of Open Source Software* **2019**, *4*, 1831.

(90)    Cuchillo, R.; Pinto-Gil, K.; Michel, J. A Collective Variable for the Rapid Exploration of Protein Druggability. *Journal of Chemical Theory and Computation* **2015**, *11*, 1292-1307.

(91)    Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *Bmc Bioinformatics* **2009**, *10*.

(92)    Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tuffery, P. fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Research* **2010**, *38*, W582-W589.

(93)    Kuzmanic, A.; Bowman, G. R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Accounts of Chemical Research* **2020**, *in press*.

## GRAPHICAL TOC