Exhaustive Search of Ligand Binding Pathways via Volume-based Metadynamics

Riccardo Capelli,[*,†,‡] Paolo Carloni,[†,¶] and Michele Parrinello[§,‖,⊥]

†*Computational Biomedicine (INM-9/IAS-5), Forschungszentrum Jülich, Wilhelm-Johnen-Straße, D-52425 Jülich, Germany*

‡*JARA-HPC, Forschungszentrum Jülich, D-54245 Jülich, Germany*

¶*Department of Physics, RWTH Aachen University, D-52078 Aachen, Germany*

§*Department of Chemistry and Applied Biosciences, ETH Zürich, c/o USI Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Ticino, Switzerland*

‖*Facoltà di Informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana (USI), Via Giuseppe Buffi 13, CH-6900, Lugano, Ticino, Switzerland*

⊥*Istituto Italiano di Tecnologia, Via Morego 30, I-16163 Genova, Italy*

E-mail: r.capelli@fz-juelich.de

Abstract

Determining the complete set of ligands' binding/unbinding pathways is important for drug discovery and to rationally interpret mutation data. Here we have developed a metadynamics-based technique that addressed this issue and allows estimating affinities in the presence of multiple escape pathways. Our approach is shown on a Lysozyme T4 variant in complex with the benzene molecule. The calculated binding free energy is in agreement with experimental data. Remarkably, not only we were able to find all the previously identified ligand binding pathways, but also we identified 3 pathways previously not identified as such. This results were obtained at a small computational

cost, making this approach valuable for practical applications, such as screening of small compounds libraries.

Describing mechanisms and energetics of ligand binding and unbinding from their targets is of great importance in drug design. The prediction of poses, affinities, and binding kinetics helps in understanding the effect of chemical decorations on the ligand and/or in mutations in the host system.[1] These important problems are most often associated with rare events and they can be adequately studied only with enhanced sampling (ES) methods, that allow estimating the free energy as a function of appropriate collective variables (CVs). In this context, we have recently used one of such methods (funnel[2] well-tempered metadynamics[3,4]), combined with novel and powerful dimension reduction approach to identify the CVs, to predict the free energy landscape of a ligand binding to a typical membrane receptor, the muscarinic M_2.[5] The picture that did emerge from our study was relatively simple, with two rather diverse escape routes of the ligand towards the extracellular region. However, proteins can exhibit a much larger flexibility with many more diverse escape pathways.[6] This is the case of the well-studied lysozyme T4 L99A variant (hereafter T4L)[7–9] (see Figure 1) in complex with benzene. The protein fold consists of two domains: the N-terminal one (residues 1-70) formed by 3 α-helices and 3 antiparallel β-strands, and a barrel-shaped C-terminal domain (residues 71-162) formed by 8 α-helices, in which the ligand accommodates. The protein features several binding/unbinding pathways, whose complete picture is not yet clear; indeed, a variety of ES approaches, with different force fields and solvent representations (see Table 2) have identified up to five pathways involving the C-terminal domain shown in Figure 4 (See SI).

Clearly, a computationally inexpensive methodology which would permit as an exhaustive as possible exploration of all the binding pathways, along with their energetics, would be highly desirable for a theory point of view and for drug discovery applications.

Here, we introduce a new practical method for addressing this issue at a modest computational cost. We shall do this in the frame of Well-tempered Metadynamics (WT-MetaD).

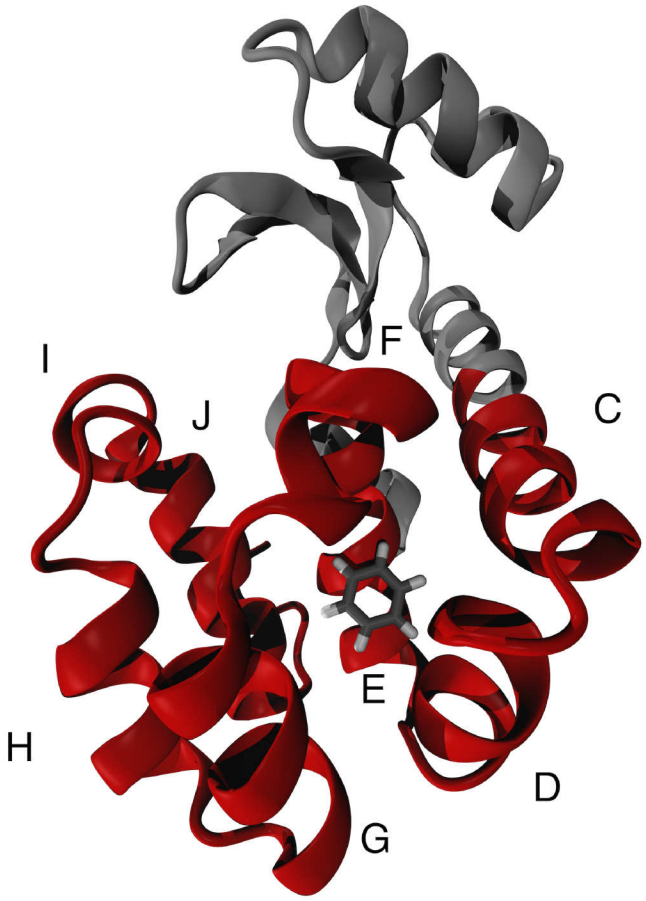

Figure 1: Cartoon representation of the T4L protein. The C-terminal domain is depicted in red color. The benzene ligand is represented in dark grey licorice. The N-terminal domain, not involved in the binding process, is in light grey. The helices of the C-terminal domain from C to H are labeled following Rydzewski et al.[10]

.

As it is well known, WT-MetaD relies on the choice of an appropriate set of CVs. We take advantage of the fact that T4L flexibility is limited and thus its center of mass can be defined once and for all. Centered around this position we consider a sphere of finite radius ρ_s, larger than the radius of gyration of our protein. Within this volume we use as CVs the spherical coordinates (ρ, θ, φ). A repulsive potential is added at the border of the sphere so as to limit the volume in the solvated state that needs to be sampled, increasing the probability of a

subsequent recrossing event. The restraining potential is in the form

$$U_{\mathrm{s}}(\rho(t)) = \begin{cases} \frac{1}{2}k\left(\rho(t) - \rho_{\mathrm{s}}\right)^2 & \text{if } \rho(t) > \rho_{\mathrm{s}} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where k has to be large enough to prevent the ligand escaping from the confining volume. In equation (1) $\rho(t)$ is the distance of the ligand from the center of mass of the target protein, and ρ_{s} is the radius of the spherical restraint.

This approach is an extension of funnel metadynamics, where a funnel-shaped potential limits the volume accessible to the ligand in the solvated space, considering a single port of entry for the ligand. Here, we are instead able to observe any binding/unbinding pathway accessible.

The application of such a restraining potential causes a change in the translational entropy of the solvated states. With the same spirit of previous works,[2,11] the correction to the binding free energy due to this constraint can be estimated to be

$$\Delta G_b^0 = -RT \log\left(C^0 K_b\right) = \Delta G_{\mathrm{MetaD}} - RT \log\left(\frac{V^0}{\frac{4}{3}\pi\rho_{\mathrm{s}}^3 - V_{\mathrm{prot}}}\right) \tag{2}$$

where R is the gas constant, T is the system temperature, K_b is the binding constant, $C^0 = 1660\text{Å}^3$ is the standard concentration, V^0 is its reciprocal, $\Delta G_{\mathrm{MetaD}}$ is the binding free energy obtained by WT-MetaD, and V_{prot} is the volume of the protein inside the restraining potential. The derivation of this correction can be found on the SI. The rotational entropy contribution is fully taken into account in our framework, because (i) the presence of the spherical restraint does not act on the rotational degrees of freedom of the ligand, and (ii) the metadynamics approach does not affect the rotational entropy.

To build our model, we used an experimental X-ray structure of T4L /benzene complex (PDB code: 1L84[7]). Details on the system preparation, equilibration and run parameters

can be found in the SI. The simulation were performed with GROMACS 2018.3,[12] patched with Plumed 2.5.[13]

We performed 5 different 200-ns long WT-MetaD simulations, using different values of ρ_s. After reaching convergence, the calculated free energy surface (FES) as a function of ρ, θ, φ is projected (with a reweighting procedure[14]) on two more informative CVs. These are the distance from the center of mass of the C-terminal domain ρ and a CV that measures whether the ligand is in contact with the host protein, measured by the coordination number C_N

$$C_N = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{\mathbf{r}_{ij}}{r_0}\right)^n}{1 - \left(\frac{\mathbf{r}_{ij}}{r_0}\right)^m} \tag{3}$$

where A is the set of non-hydrogen atoms of the ligand, B is the set of non-hydrogen atoms of the protein, r_0=4.5 Å ,is the threshold to define a formed contact, r_{ij} is the distance between atoms i and j, and $m = 6$ and $n = 12$ are the exponents of the switching function. The projected FES (Figure 2) is shown as a function of different ρ_s in order to study the influence of this parameter on the results. It is evident that, especially for shorter values of ρ_s, the presence of the restraining potential induces some artificial minima at the border of the ligand accessible volume. However, these artifacts decrease with increasing the ρ_s value and in our case the estimate of ΔG_b^0 does not change in a sensible way with ρ_s, appearing to be consistent with the experimental data (see Table 1) evaluation. $\Delta G_{\mathrm{MetaD}}$ was estimated as the ratio of the sum of Boltzmann weights extracted from the reweighted free energy surfaces. The statistical error on the value of $\Delta G_{\mathrm{MetaD}}$ (and thus of ΔG_b^0) was obtained by block analysis technique.

We defined the bound state of our system as the area of the reweighted free energy surface in the interval $\rho \in [4\text{Å}, 6\text{Å}]$ and $C_N \in [80, 120]$ (see Figure 3). The unbound state can be selected arbitrarily keeping in mind that it should be associated with as little weak non-bonded protein/ligand interactions as possible. Here, we identified this state as the region

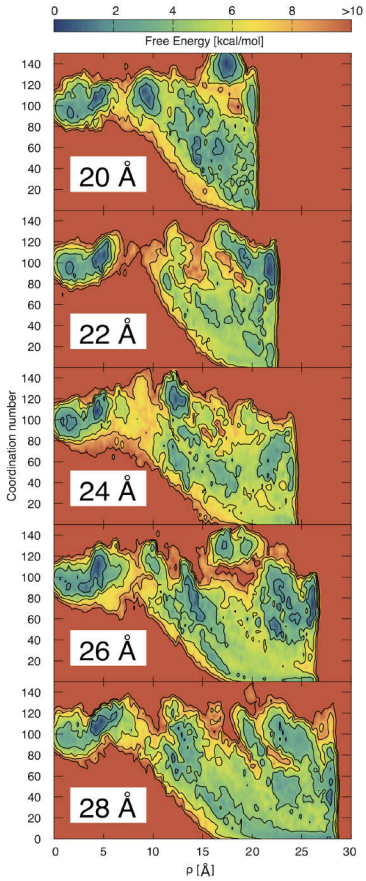

Figure 2: Free energy surfaces in function of the restraint radius ρ_s.

of the free energy surface for which the protein-ligand distance is largest (> 19Å) and with the lowest coordination number (area of the free energy surface in the interval $\rho \in [19\text{Å}, \rho_s]$ and $C_N < 10$, see Figure 3). In this region, the ligand is fully hydrated and the enthalpic contribution of ligand/protein interactions contribute negligibly to the free energy (data not shown).

Having validated our approach, we turn our attention to the pathways emerging from the WT-MetaD calculations. These are extracted by visual inspection of the trajectories. In this work, we discriminate the different unbinding pathways observing the helices that interacts in the transition between the bound and the solvated state. With this classification we identified eight different pathways (**A**-**H** in Figure 4). Of those eight, five (**A**-**E**) have

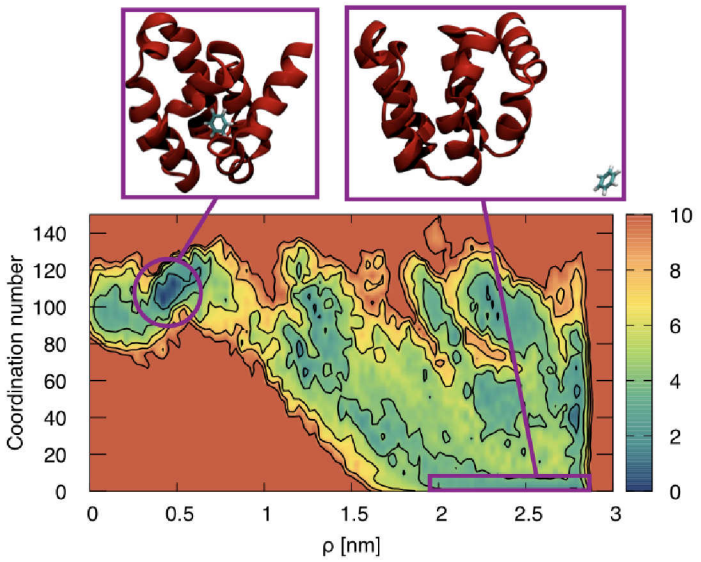

Figure 3: Free energy surface for $\rho_s = 28$ Å. We highlighted the position of the bound and the unbound state in the free energy surface, with an example of the 3D conformations involved.

Table 1: Absolute free energy differences obtained with our technique at different radii of the restraining sphere compared with experimental data.

ρ_s [Å]	ΔG_{MetaD} [kcal/mol]	$T\Delta S$ [kcal/mol]	ΔG_b^0 [kcal/mol]
20	-3.6 ± 0.5	1.6	-5.2 ± 0.5
22	-3.2 ± 0.4	1.8	-5.0 ± 0.4
24	-3.8 ± 0.4	2.0	-5.8 ± 0.4
26	-3.4 ± 0.5	2.2	-5.6 ± 0.5
28	-3.1 ± 0.4	2.3	-5.4 ± 0.4
Experimental[15]	–	–	-5.2 ± 0.2

already been reported in the literature as different unbinding pathways[10,16] (see Table 2)[1].

We discuss here only the new found ones (for a full description of all the 8 pathways, see the SI). In pathway **F** the ligand moves toward the helices H, I, and J (see Figure 1). In contrast with pathway **C**, that shares this first phase, during the solvation phase the ligand does not interact with the helices G and H, and passes through the helices F and I. Pathway **G** involve the passage of the ligand between the 3 helices H, I and J, slightly broadening the distance between helices H and J.

[1]After the submission of the present manuscript, Rydzewski and Valsson added a note in a new version of their preprint (v4 on arXiv) stating that the pathways **C**, **F**, **G**, and **H** are subclasses of the pathway identified by pwc by them.[10] No details of this claim were provided.

Table 2: Comparison of the different techniques with respective results on the number of identified pathways. The published papers by Miao et al., Wang et al., Nunes-Alves et al., and the manuscript by Rydzewski and Valsson available on the internet are reported.

Source	CV-based	Force Field	Water	# Pathways
Miao et al.[17]	No	Amber99SB	explicit	1
Wang et al.[18]	Yes	CHARMM22*	explicit	1
Nunes-Alves et al.[16]	No	CHARMM36	implicit	4
Rydzewski et al.[10]	Yes	OPLS-AA/L	explicit	5
This work	Yes	Amber14SB	explicit	8

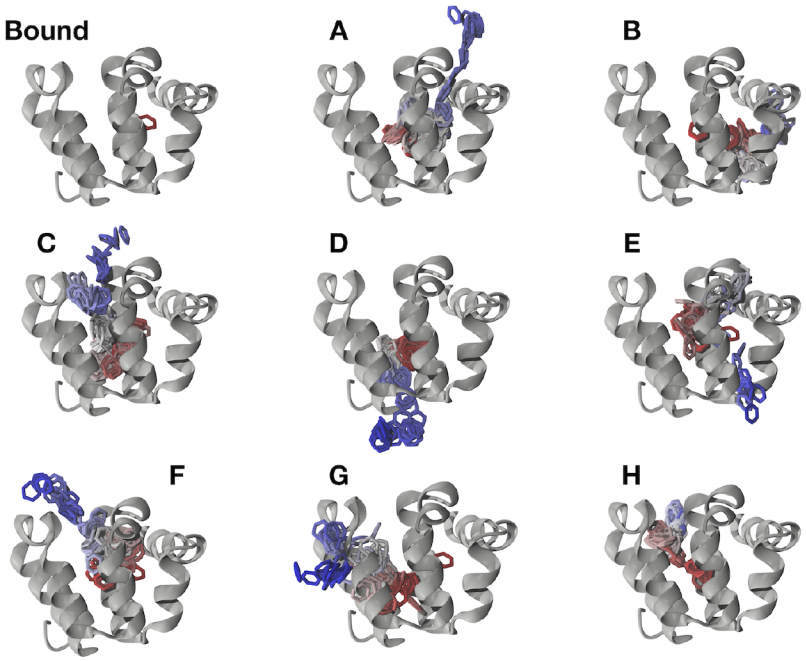

Figure 4: Bound state and binding pathways (**A** to **H**) of benzene to T4L C-terminal domain (gray). The benzene ligand is colored from red to blue on passing from the bound to the unbound state.

In contrast to all the other pathways, **H** is a one-way pathway (see Figure 5): the ligand enters a region between helices E and H, but it does not follow the same route during the unbinding process. This is most probably hampered by VAL149 and MET102 side chains, which sterically hinder the ligand only in the unbinding direction. So, we always observe binding events for this pathway. Indeed, every time we saw a transition that follow this route the ligand exited from one of the other 7 pathways. Interestingly, the single-point mutation M102Q has been studied previous in the literature to favor the binding of polar ligand inside

T4L.[9,19]

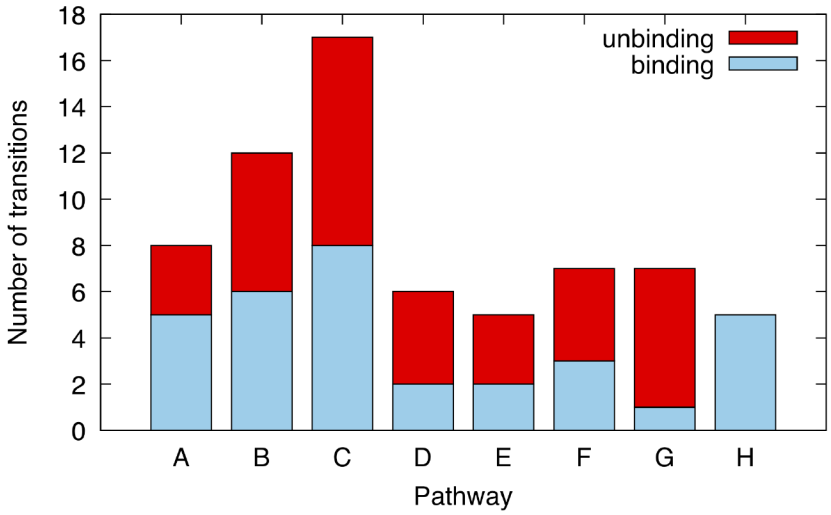

Figure 5: Stacked histogram of the observed binding/unbinding transitions via different pathways. The data is coming from all the 5 metadynamics simulations with different ρ_s values (see equation 2). The unbinding and binding processes are in red and blue, respectively.

In conclusion, we presented a highly efficient approach to sample all the possible binding/unbinding pathways for a ligand inside a host protein. We shown that the choice of the potential radius ρ_s does not influence the agreement with experimental data of our ΔG_b^0 estimation after the application of the entropic correction. Our approach was also able to identify all the 5 previously found binding pathway, and to identify 3 new pathways (**F**, **G**, and **H**).

Our technique is totally general and does not need any information regarding the pathways and the bound state. Despite our analysis of the pathways is only qualitative (i.e. we did not reweight the probability of every pathway with respect to the bias), our results can be used to build ad-hoc CVs, like pathCV,[20] to obtain the free energy profile of the binding/unbinding transition for every pathway.

All the data and PLUMED input files required to reproduce the results reported in this paper are available on PLUMED-NEST (www.plumed-nest.org), the public repository of the

PLUMED consortium,[21] as plumID:19.017.

Acknowledgement

The authors want to thank GiovanniMaria Piccini, Michele Invernizzi, Loris Di Cairano, Luca Maggi, Mauro Pastore, and Jakub Rydzewski for fruitful discussion. The authors gratefully acknowledge the computing time granted through VSR on the supercomputer JURECA[22] at Forschungszentrum Jülich (project ID: jias5d). This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 785907 (HBP SGA2).

Supporting Information Available

Preparation and simulation details, derivation of the entropic correction, convergence of ΔG_b^0 estimates with different values of ρ_s, a schematic representation of the volume bias, all pathways description, and PLUMED input files are available in the Supporting Information.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Held, M.; Metzner, P.; Prinz, J.-H.; Noé, F. Mechanisms of protein-ligand association and its modulation by protein mutations. *Biophys. J.* **2011**, *100*, 701–710.

(2) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358–6363.

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

(4) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(5) Capelli, R.; Bochicchio, A.; Piccini, G.; Casasnovas, R.; Carloni, P.; Parrinello, M. Chasing the full free energy landscape of neuroreceptor/ligand unbinding by metadynamics simulations. *J. Chem. Theory Comput.* **2019**,

(6) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. Protein binding pocket dynamics. *Acc. Chem. Res.* **2016**, *49*, 809–815.

(7) Eriksson, A.; Baase, W.; Wozniak, J.; Matthews, B. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **1992**, *355*, 371.

(8) Morton, A.; Matthews, B. W. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry* **1995**, *34*, 8576–8588.

(9) Baase, W. A.; Liu, L.; Tronrud, D. E.; Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **2010**, *19*, 631–641.

(10) Rydzewski, J.; Valsson, O. Finding Multiple Reaction Pathways of Ligand Unbinding. *arXiv preprint arXiv:1808.08089* **2018**,

(11) Allen, T. W.; Andersen, O. S.; Roux, B. Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 117–122.

(12) 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*, 19–25.

(13) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(14) Tiwary, P.; Parrinello, M. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2014**, *119*, 736–742.

(15) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry* **1995**, *34*, 8564–8575.

(16) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M. Escape of a small molecule from inside T4 lysozyme by multiple pathways. *Biophys. J.* **2018**, *114*, 1058–1066.

(17) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584–3595.

(18) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chem. Sci.* **2017**, *8*, 6466–6473.

(19) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for docking. *J. Med. Chem.* **2005**, *48*, 3714–3728.

(20) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.

(21) The PLUMED consortium, The PLUMED consortium: A community effort to promote openness, transparence and reproducibility in molecular simulations. *Submitted* **2019**,

(22) Jülich Supercomputing Centre, JURECA: Modular supercomputer at Jülich Supercomputing Centre. *Journal of large-scale research facilities* **2018**, *4*.

Graphical TOC Entry